December 2019

Copula-based Multimodal Data Fusion for Inference with Dependent Observations

Shan Zhang
Syracuse University

ABSTRACT

Fusing heterogeneous data from multiple modalities for inference problems has been an attractive and important topic in recent years. There are several challenges in multi-modal fusion, such as data heterogeneity and data correlation. In this dissertation, we investigate inference problems with heterogeneous modalities by taking into account nonlinear cross-modal dependence. We apply copula based methodology to characterize this dependence.

In distributed detection, the goal often is to minimize the probability of detection error at the fusion center (FC) based on a fixed number of observations collected by the sensors. We design optimal detection algorithms at the FC using a regular vine copula based fusion rule. Regular vine copula is an extremely flexible and powerful graphical model used to characterize complex dependence among multiple modalities. The proposed approaches are theoretically justified and are computationally efficient for sensor networks with a large number of sensors.

With heterogeneous streaming data, the fusion methods applied for processing data streams should be fast enough to keep up with the high arrival rates of incoming data, and meanwhile provide solutions for inference problems (detection, classification, or estimation) with high accuracy. We propose a novel parallel platform, C-Storm (Copula-based Storm), by marrying copula-based dependence modeling for highly accurate inference and a highly-regarded parallel computing platform Storm for fast stream data processing. The efficacy of C-Storm is demonstrated.

In this thesis, we consider not only decision level fusion but also fusion with heterogeneous high-level features. We investigate a supervised classification problem by fusing dependent high-level features extracted from multiple deep neural network (DNN) classifiers. We employ regular vine copula to fuse these high-level features. The efficacy of the combination of model-based method and deep learning is demonstrated.

Besides *fixed-sample-size* (FSS) based inference problems, we study a distributed sequential detection problem with *random-sample-size*. The aim of the distributed sequential detection problem in a non-Bayesian framework is to minimize the average detection time while satisfying the pre-specified constraints on probabilities of false alarm and miss detection. We design local memory-less truncated sequential tests and propose a copula based sequential test at the FC. We show that by suitably designing the local thresholds and the truncation window, the local probabilities of false alarm and miss detection of the proposed local decision rules satisfy the pre-specified error probabilities. Also, we show the asymptotic optimality and time efficiency of the proposed distributed sequential scheme.

In large scale sensors networks, we consider a collaborative distributed estimation problem with statistically dependent sensor observations, where there is no FC. To achieve greater sensor transmission and estimation efficiencies, we propose a two-step cluster-based collaborative distributed estimation scheme. In the first step, sensors form dependence driven clusters such that sensors in the same cluster are dependent while sensors from different clusters are independent, and perform copula-based maximum a posteriori probability (MAP) estimation via intra-cluster collaboration. In the second step, the estimates generated in the first step are shared via inter-cluster collaboration to reach an average consensus. The efficacy of the proposed scheme is justified.

COPULA-BASED MULTIMODAL DATA FUSION FOR INFERENCE WITH DEPENDENT OBSERVATIONS

By

Shan Zhang
B.S., University of Science & Technology Beijing, China, 2014

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical and Computer Engineering

Syracuse University
December 2019

Acknowledgments

First and foremost, I would like to express my deepest gratitude and sincere thanks to my advisor, Prof. Pramod K. Varshney, for his guidance, support and encouragement. I have been amazingly fortunate to have him during my doctoral study. He gave me the freedom to explore on my own, and at the same time guidance to recover when my steps faltered. He has been influencing me with his enthusiasm towards research as well as life, and will keep guiding me in the future.

Besides my advisor, I would like to thank Mrs. Varshney for her kindness, support and all the home-cooked meals. She is like a mother to me. She is a very strong and inspiring woman. She made me understand the true meaning of love. I have learned a lot from her, and will miss her very much.

I would like to thank the rest of my thesis committee: Prof. Biao Chen, Prof. Lixin Shen, Prof. Mustafa C. Gursoy, Prof. Chilukuri K. Mohan and Prof. Reza Zafarani for their insightful comments and suggestions. Dr. Lakshmi N. Theagarajan and Dr. Sora Choi have provided technical support for part of the work in this thesis.

Also, I would like to thank all my fellow labmates in the Sensor Fusion Laboratory for their constant support in my research and life: Pranay Sharma, Prashant Khanduri, Swatantra Kafle, Saikiran Bulusu, Qunwei Li, Baocheng Geng, Quan Chen and Nandan Sriranga. Also, I would like to take this opportunity to thank Dr. Sijia Liu, Dr. Aditya Vempaty, Prof. Vinod Sharma and Prof. Jian Tang for their patient guidance and support. Especially, I am very grateful to have Dr. Nianxia Cao, Dr. Hao He, Dr. Fangrong Peng, my friends Mingyue Wang and Qianyun Wang during my doctoral

studies. Their unreserved support means a lot to me.

Most importantly, none of this would have been possible without the love and constant encouragement of my parents, my brother and my husband. I would like to thank them for always being there cheering me up and standing by me through the good times and bad.

TABLE OF CONTENTS

Acknowledgments **v**

List of Tables **xi**

List of Figures **xiii**

1 Introduction **1**

1.1 Background 2

1.1.1 Copula Theory 3

1.1.2 Summary of Some Multivariate Copula Functions 4

1.1.3 Copulas and Measures of Dependence 5

1.1.4 Regular Vine Copula 6

1.1.5 Array Representation of Regular Vine 10

1.2 Literature Review 11

1.2.1 Linear Dependence: Covariance Matrix 12

1.2.2 Nonlinear Dependence: Nonparametric Approach 13

1.2.3 Nonlinear Dependence: Copula-based Approach 14

1.3 Main Contributions and Organization 16

1.4 Bibliographic Note 18

2 Distributed Detection Based on Regular Vine Copulas **20**

2.1 Motivation 20

2.2 Problem Formulation 21
2.3 R-Vine Copula Based Fusion of Multiple Statistically Dependent Decisions . . 24
2.3.1 Optimal Test Statistic 24
2.3.2 R-Vine Copula Based Dependence Modeling 28
2.3.3 Model Selection and Estimation 29
2.4 Efficient R-Vine Copula Based Fusion with Statistical Dependent Decisions . . 31
2.5 Simulation Results 34
2.6 Summary 42

3 Copula based Distributed Parallel Computing Platform 43
3.1 Motivation 43
3.2 Storm 44
3.3 Design of C-Storm 45
3.3.1 Architecture of C-Storm 45
3.3.2 Copula-based Dependence Modeling 48
3.4 Simulation Results 50
3.4.1 Fusion Application and Experimental Setup 50
3.4.2 Experimental Results and Analysis 52
3.5 Summary 55

4 Distributed Classification with Dependent Features 56
4.1 Motivation 56
4.2 Problem Formulation 58
4.3 R-Vine Copula Based Fusion of Multiple Deep Neural Networks 60
4.3.1 R-Vine copula Models 61
4.3.2 Estimation of Optimal R-Vine copula 61
4.4 Simulation Results 62

4.4.1 Datasets . . . 63
4.4.2 Classification Accuracy . . . 64
4.5 Summary . . . 66

5 Distributed Sequential Detection with Dependent Observations 68
5.1 Motivation . . . 68
5.2 Problem Formulation . . . 69
5.3 Centralized Copula-based Sequential Probability Ratio Test . . . 72
5.4 Distributed Copula-based Sequential Probability Ratio Test . . . 77
5.4.1 Local Sensor Detection Rule . . . 78
5.4.2 Derivation of the Fusion Rule . . . 81
5.5 Simulation Results . . . 83
5.6 Summary . . . 90

6 Distributed Estimation in Large Scale Wireless Sensor Networks via A Two-Step Cluster-based Approach 92
6.1 Motivation . . . 92
6.2 Problem Formulation . . . 93
6.3 Two-Step Dependence Driven Collaborative Distributed Estimation . . . 97
6.3.1 Assumptions and Dissimilarity Measure Definitions . . . 98
6.3.2 Dependence Driven Clustering Process . . . 99
6.3.3 Copula Based MAP . . . 100
6.3.4 Cluster Based Consensus Scheme . . . 103
6.4 Sensor Selection Based Two-Step Dependence Driven Collaborative Distributed Estimation . . . 105
6.5 Simulation Results . . . 108
6.6 Summary . . . 116

7 Summary and Future Directions **118**

7.1 Summary 118

7.2 Future Directions 121

Appendix **123**

A $A_{\mathbf{u}^t}$ in Log Test Statictics (2.6) 123

B Proof of Theorem 2.1 124

C Proof of Theorem 5.1 125

D Proof of Theorem 5.2 128

E Proof of Theorem 5.3 129

F Proof of Theorem 6.1 130

G Proof of Theorem 6.2 130

References **135**

LIST OF TABLES

1.1 Archimedean copula functions. 5

2.1 The performance of R-Vine classes and standard multivariate copulas. 37

4.1 **STISEN**: F_1 scores for Watch-DNN, Phone-DNN, Fully-connected layer fusion, Data-level fusion, R-Vine copula fusion. 64

4.2 **ANGUITA**: F_1 scores for Accelerometer-DNN, Gyroscope-DNN, Fully-connected layer fusion, Data-level fusion, R-Vine copula fusion. 64

4.3 **STISEN**: Confusion matrix for R-Vine copula based fusion. 66

4.4 **ANGUITA**: Confusion matrix for R-Vine copula based fusion. 66

5.1 Known copula: Estimated P_F and P_M with $\alpha = 0.01$, $L = 3$ for centralized sequential scheme and Case 1. 85

5.2 Known copula: Estimated P_F and P_M with $\beta = 0.01$, $L = 3$ for centralized sequential scheme and Case 1. 85

5.3 Average p values for the estimation of underlying dependence using R-Vine copula model with different number of sensors. 86

5.4 Average p values for the estimation of underlying dependence using different multivariate copula models with $L = 3$. 87

5.5 Unknown copula: Estimated P_F, P_M and $\mathbb{E}[T]$ with $\alpha = \beta = 0.01$ and SNR $= -6$ dB. 89

5.6 Unknown copula: Estimated P_F, P_M and $\mathbb{E}[T]$ with $\alpha = \beta = 0.01$ and SNR $=$ -9 dB. 89
5.7 Unknown copula: Estimated P_F, P_M and $\mathbb{E}[T]$ with $\alpha = \beta = 0.01$ and SNR $=$ 0 dB for Case 2. 90

LIST OF FIGURES

1.1 An example R-Vine for five variables. 8

2.1 ROCs comparing the Chair-Varshney fusion rule and the R-Vine copula based fusion rule with dependent fading channels. 39
2.2 ROCs comparing the Chair-Varshney fusion rule and the R-Vine copula based fusion rule with dependent signals. 39
2.3 ROCs comparing the Chair-Varshney fusion rule and the R-Vine copula based fusion rule with dependent signals for weaker dependence. 40
2.4 ROCs comparing the Chair-Varshney fusion rule and the R-Vine copula based fusion rule with dependent signals for $q_l \leq 0.05$. 40
2.5 ROCs for R-Vine copula based fusion rule with dependent signals for three model selection criteria. 41
2.6 ROCs comparing the Chair-Varshney fusion rule and the R-Vine copula based fusion rule with dependent measurement noise. 42

3.1 The architecture of Storm. 44
3.2 The architecture of C-Storm. 46
3.3 C-Storm versus sequential baseline in terms of average total processing time. . 52
3.4 Average total processing times of C-Storm with different number of workers used for obtaining the marginal probability densities in the fusion bolt. 53

3.5 Average total processing times of C-Storm with different number of workers used for performing the test in the fusion bolt. 54
3.6 ROCs for the detection problem. 54

4.1 A typical Deep Neural Network structure [84]. 58
4.2 R-Vine copula based multi-modal DNN. 58
4.3 First level dependence structure for activity 'Walking-upstairs'. 65

5.1 Parallel distributed detection system. 70
5.2 Average expected stopping time as a function of α for centralized sequential scheme. 86
5.3 Average expected stopping time as a function of β for centralized sequential scheme. 86
5.4 Truncation window W_0 as a function of $\tilde{\alpha}$ and $\tilde{\beta}$, where $\tilde{\alpha} = \tilde{\beta}$. 88

6.1 Two-step cluster-based collaborative distributed estimation system, where the orange dash lines represent the inter-cluster communication links and the black dash lines denote the intra-cluster communication links. 94
6.2 Average clustering accuracy as a function of threshold d_{th}. 110
6.3 Average clustering accuracy as a function of number of observations N with $d_{th} = 0.83$. 110
6.4 Average MSE as a function of SNR without sensor selection. 111
6.5 Average MSE as a function of the number of observations N without sensor selection. 111
6.6 Average MSE as a function of SNR with cluster-based sensor selection and $m_k = 3$. 112
6.7 Average MSE as a function of the number of observations N with cluster-based sensor selection and $m_k = 3$. 113

6.8 Average MSE as a function of SNR for different schemes without sensor selection. 114
6.9 Average MSE as a function of the number of observations N for different schemes without sensor selection. 114
6.10 Average MSE as a function of SNR for the cluster-based sensor selection scheme and the global sensor selection scheme. 115
6.11 Average MSE as a function of the number of observations N for the cluster-based sensor selection scheme and the global sensor selection scheme. 116

CHAPTER 1
INTRODUCTION

The problem of inference by fusing data from multiple modalities has a wide variety of applications, such as activity monitoring, autonomous robotics and military/security surveillance. Typically, a large number of spatially distributed sensors are deployed in a network and these sensors operate collaboratively to solve an inference problem, such as detection, estimation and classification. Fusing observations of multiple sensors can improve decision making and provide global information of a certain phenomenon. However, sensors used for observing the same phenomenon are usually of different modalities, namely, they are incommensurate/heterogeneous. Sensors are said to be heterogeneous if their respective observation models cannot be described by the same statistical distribution. Moreover, sensor observations are often dependent due to a variety of reasons such as sensing of the same phenomenon. The nature of this dependence can be quite complex and nonlinear, especially in cases where the signal may propagate through a non-homogeneous medium. Inference in such multi-sensor systems is the major topic of this thesis.

In networks with limited communication resources, local observations are usually compressed at the sensors according to certain local rules, and only the compressed information is transmitted to the FC. In such distributed networks, the challenge is to achieve high perfor-

mance in terms of accuracy efficiency and time efficiency while satisfying energy and bandwidth constraints. The existence of nonlinear cross-modal dependence and heterogeneity of sensors in the network make the design of local inference rules and the fusion rule at the FC highly complex. In terms of accuracy, we study the design of local and fusion rules in this thesis, where we take the underlying spatial dependence into consideration to improve inference performance. In terms of time efficiency, we consider a distributed sequential network, and design sequential tests at the local sensors and a copula based sequential test at the FC. A parallel platform for fusing heterogenous streaming data is also investigated to accelerate inference response. Moreover, in a fully distributed network with no FC, intra-cluster collaboration and inter-cluster collaboration are studied to exploit the underlying dependence among sensors so that inference performance is improved to the largest extent under limited communication budget.

1.1 Background

Copula theory, which forms the basis of a lot of work in this thesis, is presented in this section. Copulas provide a flexible and powerful approach for modeling underlying dependence among continuous random variables. A multivariate copula, specified independently from marginals, is a multivariate distribution with uniform marginal distributions. The unique correspondence between a multivariate copula and any multivariate distribution is stated in Sklar's theorem [75] which is a fundamental theorem of copula theory. Standard well defined multivariate copulas may lack the ability to model high dimensional nonlinear dependencies due to factors such as limited number of parameters to characterize the dependence. Based on this, regular vine copula based methodology has been developed for more flexible modeling of dependencies in larger dimensions. In the following, we first give the theoretical background of copula theory and present some well defined multivariate copulas, and then introduce the regular vine copula.

1.1.1 Copula Theory

Theorem 1.1 (Sklar's Theorem)**.** *The joint distribution function F of random variables $x_1, \ldots, x_d$ with continuous marginal distribution functions $F_1, \ldots, F_d$ can be cast as*

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)|\boldsymbol{\phi}), \tag{1.1}$$

where C is a unique d-dimensional copula with dependence parameter $\boldsymbol{\phi}$. Conversely, given a copula C and univariate Cumulative Distribution Functions (CDFs) $F_1, \ldots, F_d$, F in Equation (1.1) *is a valid multivariate CDF with marginals $F_1, \ldots, F_d$. Note that $\boldsymbol{\phi}$ is used to characterize the amount of dependence among the d random variables. In general, $\boldsymbol{\phi}$ may be a scalar, a vector or a matrix.*

For absolutely continuous distributions F and $F_1, \ldots, F_d$, the joint Probability Density Function (PDF) of random variables $x_1, \ldots, x_d$ can be obtained by differentiating both sides of Equation (1.1):

$$f(x_1, \ldots, x_d) = \Big(\prod_{m=1}^{d} f_m(x_m) \Big) c(F_1(x_1), \ldots, F_d(x_d)|\boldsymbol{\phi}), \tag{1.2}$$

where $f_1, \ldots, f_d$ are the marginal densities and c is referred to as the density of multivariate copula C that is given by

$$c(\mathbf{u}|\boldsymbol{\phi}) = \frac{\partial^d (C(u_1, \ldots, u_d|\boldsymbol{\phi}))}{\partial u_1, \ldots, \partial u_d}, \tag{1.3}$$

where $u_m = F_m(x_m)$ and $\mathbf{u} = [u_1, \ldots, u_d]$.

Thus, given specified univariate marginal distributions $F_1, \ldots, F_d$ and the copula model C, the joint distribution function F can be constructed by

$$F(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_d^{-1}(u_d)) = C(u_1, u_2, \ldots, u_d|\boldsymbol{\phi}), \tag{1.4}$$

where $u_m = F_m(x_m)$ and $F_m^{-1}(u_m)$ are the inverse distribution functions of the marginals, $m = 1, 2, \ldots, d$.

Note that $C(\cdot|\boldsymbol{\phi})$ is a valid CDF and $c(\cdot|\boldsymbol{\phi})$ is a valid PDF for uniformly distributed random variables u_m, $m = 1, 2, \ldots, d$. Since different copula functions may model different types of dependence, selection of copula functions to characterize joint statistics of random variables is a key problem. Various families of multivariate copula functions are described in [75]. A brief summary of some popularly used copula functions is provided next.

1.1.2 Summary of Some Multivariate Copula Functions

Elliptical copulas

The Gaussian and the Student-t copula functions belong to the family of elliptical copulas. They are derived from multivariate Gaussian and Student-t distributions, respectively. They both specify dependence using the correlation matrix and are given as follows.

The multivariate Gaussian copula, derived from a d-dimensional multivariate Gaussian distribution, is defined as

$$C^G(\mathbf{u}|\boldsymbol{\Sigma}) = \Phi_{\boldsymbol{\Sigma}}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)), \tag{1.5}$$

where $\boldsymbol{\Sigma}$ is the correlation matrix, Φ is the univariate normal CDF and $\Phi_{\boldsymbol{\Sigma}}$ denotes the multivariate normal CDF.

Similarly, the Student-t copula is derived from a d-dimensional multivariate Student-t distribution, which is given by

$$C^t(\mathbf{u}|\boldsymbol{\Sigma}, \nu) = t_{\nu,\boldsymbol{\Sigma}}(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)), \tag{1.6}$$

where $t_{\nu,\boldsymbol{\Sigma}}$ denotes the multivariate Student-t distribution with correlation matrix $\boldsymbol{\Sigma}$ and de-

grees of freedom ν ($\nu \geq 3$), and t_ν is the univariate Student-t distribution with degrees of freedom ν. It is common to set $\nu = 3$ to incorporate heavy tail dependence. As $\nu \to \infty$, the Student-t copula approaches the Gaussian copula function.

Archimedean copulas

Archimedean copulas are defined as follows,

$$C(\mathbf{u}|\boldsymbol{\phi}) = \Psi^{-1}\left(\sum_{m=1}^{d} \Psi(u_m)\right), \tag{1.7}$$

where we refer to $\Psi(\cdot)$ as the generator function and $\boldsymbol{\phi}$ as the parameter of the copula. Some Archimedean copula functions are indicated in Table 1.1 [41].

Table 1.1: Archimedean copula functions.

Copula	Generator Function Ψ	Copulas in the Parametric Form
Clayton	$\frac{1}{\phi}\left(u^{-\phi}-1\right)$	$\left(\sum_{m=1}^{d} u_m^{-\phi}-1\right)^{-\frac{1}{\phi}},\ \ \phi \in [-1,\infty)\backslash\{0\}$
Frank	$-\log\frac{\exp\{-\phi u\}-1}{\exp\{-\phi\}-1}$	$-\frac{1}{\phi}\log\left(1+\frac{\prod_{m=1}^{d}[\exp\{-\phi u_m\}-1]}{\exp\{-\phi\}-1}\right), \phi \in \mathbb{R}\backslash\{0\}$
Gumbel	$-\ln u^{\phi}$	$\exp\left\{-\left(\sum_{m=1}^{d}(-\ln u_m)^{\phi}\right)^{\frac{1}{\phi}}\right\},\ \ \phi \in [1,\infty)$
Independent	$-\ln u$	$\prod_{m=1}^{d} u_m$

1.1.3 Copulas and Measures of Dependence

An attractive feature of copulas is that nonparametric rank-based measures of dependence, such as Kendall's τ, can be expressed as expectations over the copula distribution. For independent pairs of random variables (X_1, Y_1) and (X_2, Y_2) having the same distribution as (X, Y), concordance is defined as the condition that $(X_1 - X_2)(Y_1 - Y_2) \geq 0$ and discordance is defined as the condition that $(X_1 - X_2)(Y_1 - Y_2) < 0$. Kendall's τ is defined to be the difference between

the probabilities of concordance and discordance:

$$\tau \triangleq P[(X_1 - X_2)(Y_1 - Y_2) \geq 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Nelsen has proved the relationship in Equation (1.8) for a copula, $C(\cdot|\boldsymbol{\phi})$, and random variables $X \sim f_X(x), Y \sim f_Y(y)$ [75], i.e.,

$$\tau(\boldsymbol{\phi}) = 4 \int \int_{[0,1]^2} C(F_X(x), F_Y(y)|\boldsymbol{\phi}) dC(F_X(x), F_Y(y)|\boldsymbol{\phi}) - 1. \tag{1.8}$$

This relationship allows τ to be expressed in terms of the dependence parameter of the copula, C (Σ for the elliptical copulas and $\boldsymbol{\phi}$ for the Archimedean copulas in Table 1.1). For the case of elliptical copulas, parametrized by the matrix $\Sigma = [\rho_{ij}]$,

$$\rho_{ij} = \sin\left(\frac{\pi \tau_{ij}}{2}\right), \tag{1.9}$$

where τ_{ij} is the Kendall's τ evaluated for the pair $(U_i, U_j) = (F_{X_i}(\cdot), F_{X_j}(\cdot))$. The sample estimate of Kendall's τ, for N observations, can be calculated as the ratio of the difference in the number of concordant pairs, c_{cor}, and discordant pairs, d_{cor}, to the total number of pairs of observations, i.e.,

$$\hat{\tau} = \frac{c_{\text{cor}} - d_{\text{cor}}}{c_{\text{cor}} + d_{\text{cor}}} = \frac{c_{\text{cor}} - d_{\text{cor}}}{\binom{N}{2}}. \tag{1.10}$$

Typically, $\boldsymbol{\phi}$ is unknown a *priori* and needs to be estimated, e.g., using Maximum Likelihood Estimation (MLE) [41]. On the other hand, Equation (1.8) and Equation (1.10) imply that Kendall's τ can be used for calculating computationally efficient estimates of $\boldsymbol{\phi}$.

1.1.4 Regular Vine Copula

Regular vine (R-Vine) copulas, introduced by Bedford and Cooke in [11, 12], are extremely flexible in modeling high dimensional multivariate dependence, where a set of bivariate cop-

ulas are used to construct the multivariate copula. A regular vine copula is a tree-structured graphical model that consists of a regular vine and a set of bivariate copulas. We first present the definition of the regular vine in the following. A regular vine is defined as follows.

Definition 1.1 (R-Vine). *$\mathcal{V} = (T_1, \ldots, T_{d-1})$ is a regular vine on d elements if the following conditions are satisfied.*

1. *T_1 is a tree with nodes $N_1 = \{1, \ldots, d\}$ and a set of $d-1$ edges denoted as E_1.*
2. *For $i = 2, \ldots, d-1$, T_i is a tree with nodes $N_i = E_{i-1}$ and edge set E_i.*
3. *For $i = 2, \ldots, d-1$ and $\{a, b\} \in E_i$ with $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$, $|(a \cap b)| = 1$ (proximity condition) holds, where $|\cdot|$ denotes the cardinality of a set.*

A d-dimensional vine consists of $d(d-1)/2$ edges in total. The proximity condition implies that two edges in tree T_i are connected in tree T_{i+1} if the two edges share a common node in tree T_i.

R-Vine copula is obtained by specifying bivariate copulas, the so-called pair-copula, on each of the edges. Before introducing R-Vine copula, some sets associated with its edges need to be defined. The complete union U_e of an edge $e = \{a, b\} \in E_i, a, b \in N_i$ is defined as $U_e = \{m \in N_1 \,|\, \exists e_j \in E_j, j = 1, 2, \ldots, i-1, \text{such that} \, m \in e_1 \in \ldots e_{i-1} \in e\}$. The conditioning set of the edge $e = \{a, b\}$ is $D_e = U_a \cap U_b$ and the conditioned sets of the edge $e = \{a, b\}$ are $\mathfrak{C}_{e,a} = U_a \backslash D_e$ and $\mathfrak{C}_{e,b} = U_b \backslash D_e$. A regular vine copula is defined as follows.

Definition 1.2 (R-Vine Copula). *$(\mathbf{F}, \mathcal{V}, \mathbf{B})$ is called a R-Vine copula if*

1. *$\mathbf{F} = [F_1, F_2, \ldots, F_d]^T \in [0, 1]^d$ is a vector with uniform marginals.*
2. *$\mathcal{V}$ is a d-dimensional regular vine.*
3. *$\mathbf{B} = \{C_{\mathfrak{C}_{e,a}, \mathfrak{C}_{e,b}|D_e} \,|\, e \in E_i, i = 1, 2, \ldots, d-1\}$ is a set of bivariate copulas.*

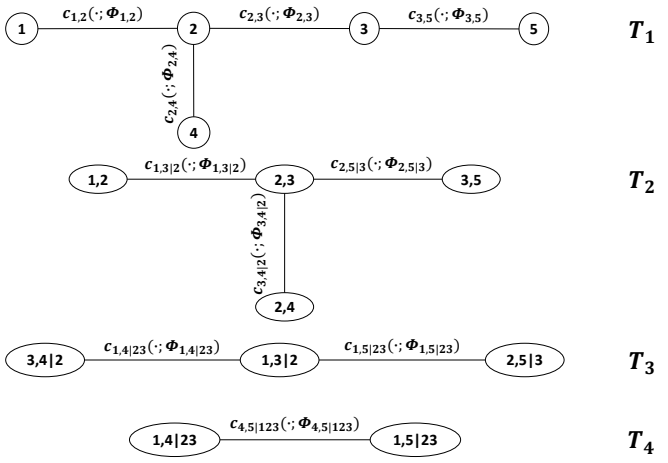

Fig. 1.1: An example R-Vine for five variables.

The joint density of a random vector $\mathbf{x} = [x_1, x_2, \ldots, x_d]^T$ is given by

$$f_{1,\ldots,d}(\mathbf{x}) = \prod_{m=1}^{d} f_m(x_m) \prod_{i=1}^{d-1} \prod_{e \in E_i} \times c_{\mathfrak{C}_{e,a},\mathfrak{C}_{e,b}|D_e}(F_{\mathfrak{C}_{e,a}|D_e}(x_{\mathfrak{C}_{e,a}}|\mathbf{x}_{D_e}), F_{\mathfrak{C}_{e,b}|D_e}(x_{\mathfrak{C}_{e,b}}|\mathbf{x}_{D_e})), \tag{1.11}$$

where $e = \{a, b\}$, $\mathbf{x}_{D_e} = \{x_j | j \in D_e\}$, f_m is the marginal PDF of variable x_m, $m = 1, \ldots, d$. The conditional distribution $F_{\mathfrak{C}_{e,a}|D_e}(x_{\mathfrak{C}_{e,a}}|\mathbf{x}_{D_e})$ can be obtained recursively tree by tree by the following equation [51].

$$F_{\mathfrak{C}_{e,a}|D_e}(x_{\mathfrak{C}_{e,a}}|\mathbf{x}_{D_e}) = \frac{\partial C_{\mathfrak{C}_{a,a_1},\mathfrak{C}_{a,a_2}|D_a}\left(F_{\mathfrak{C}_{a,a_1}|D_a}(x_{\mathfrak{C}_{a,a_1}}|\mathbf{x}_{D_a}), F_{\mathfrak{C}_{a,a_2}|D_a}(x_{\mathfrak{C}_{a,a_2}}|\mathbf{x}_{D_a})\right)}{\partial F_{\mathfrak{C}_{a,a_2}|D_a}(x_{\mathfrak{C}_{a,a_2}}|\mathbf{x}_{D_a})}, \tag{1.12}$$

where $e = \{a, b\} \in E_i$, $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$ are the edges that connect $\mathfrak{C}_{e,a}$ and $\mathfrak{C}_{e,b}$ given the conditioning variables D_e. Similarly, we can obtain $F_{\mathfrak{C}_{e,b}|D_e}(x_{\mathfrak{C}_{e,b}}|\mathbf{x}_{D_e})$.

As an example, a 5-dimensional R-Vine copula is shown in Fig. 1.1. The R-Vine has four trees T_i and the tree T_i has nodes $N_i = 6 - i$ and edges $E_i = 5 - i$, where $i = 1, 2, 3, 4$. Each edge is associated with a bivariate copula density c and its corresponding parameters $\boldsymbol{\phi}$ used to model dependence between two variables. Moreover, at each edge $e = \{a, b\} \in E_i$, the term $\mathfrak{C}_{e,a}$ and $\mathfrak{C}_{e,b}$ are separated by a comma and given to the left of the "|" sign, while

D_e appears on the right. In the first tree T_1, the dependences of the four pairs of variables $(1,2),(2,3),(2,4),(3,5)$ are modeled using four bivariate copulas, $c_{1,2}(\cdot;\boldsymbol{\phi}_{1,2})$, $c_{2,3}(\cdot;\boldsymbol{\phi}_{2,3})$, $c_{2,4}(\cdot;\boldsymbol{\phi}_{2,4})$ and $c_{3,5}(\cdot;\boldsymbol{\phi}_{3,5})$. In the second tree T_2, three conditional dependencies are modeled. The pair $(1,3|2)$ using bivariate copula density $c_{1,3|2}(\cdot;\boldsymbol{\phi}_{1,3|2})$ characterizes the dependence between the first and third variables given the second variable. Also, the pair $(3,4|2)$ using bivariate copula density $c_{3,4|2}(\cdot;\boldsymbol{\phi}_{3,4|2})$ characterizes the dependence between the third and fourth variables given the second variable. Similarly, we can obtain the bivariate copula density for the pair $(2,5|3)$. In the third tree T_3, the dependence of the first and fourth variables given the second and third variables is modeled using bivariate copula density $c_{1,4|23}(\cdot;\boldsymbol{\phi}_{1,4|23})$. Also, we can obtain the bivariate copula density for the pair $(1,5|23)$. In the fourth tree T_4, the bivariate copula density $c_{4,5|123}(\cdot;\boldsymbol{\phi}_{4,5|123})$ captures the dependence between the fourth and fifth variables given the first, second and third variables.

For the 5-dimensional case, using Equation (1.11), the joint PDF of $\mathbf{z} = [z_1, z_2, z_3, z_4, z_5]$ can be expressed as

$$
\begin{aligned}
f(z_1,z_2,z_3,z_4,z_5) = &\left[\prod_{l=1}^{5} f(z_l)\right] \cdot c_{1,2}\left(F(z_1),F(z_2);\boldsymbol{\phi}_{1,2}\right) \\
&\cdot c_{2,3}\left(F(z_2),F(z_3);\boldsymbol{\phi}_{2,3}\right) \cdot c_{2,4}\left(F(z_2),F(z_4);\boldsymbol{\phi}_{2,4}\right) \\
&\cdot c_{3,5}\left(F(z_2),F(z_3)\cdot;\boldsymbol{\phi}_{3,5}\right) \\
&\cdot c_{1,3|2}\left(F(z_1|z_2),F(z_3|z_2);\boldsymbol{\phi}_{1,3|2}\right) \\
&\cdot c_{3,4|2}\left(F(z_3|z_2),F(z_4|z_2);\boldsymbol{\phi}_{3,4|2}\right) \\
&\cdot c_{2,5|3}\left(F(z_2|z_3),F(z_5|z_3);\boldsymbol{\phi}_{2,5|3}\right) \\
&\cdot c_{1,4|23}\left(F(z_1|z_2z_3),F(z_4|z_2z_3);\boldsymbol{\phi}_{1,4|23}\right) \\
&\cdot c_{1,5|23}\left(F(z_1|z_2z_3),F(z_5|z_2z_3);\boldsymbol{\phi}_{1,5|23}\right) \\
&\cdot c_{4,5|123}\left(F(z_4|z_1z_2z_3),F(z_5|z_1z_2z_3);\boldsymbol{\phi}_{4,5|123}\right).
\end{aligned}
\tag{1.13}
$$

1.1.5 Array Representation of Regular Vine

Generally, it is quite expensive to store the nested set of trees and also not convenient to describe inference algorithms. In [72], a lower triangular array was proposed to store a R-Vine.

Definition 1.3 (R-Vine Array)**.** *A lower triangular array $M = (m_{i,j})_{i,j=1,2,\ldots,d}$ is called a R-Vine array if for $i = 1, \ldots, d-1$ and for all $k = i+1, \ldots, d-1$, there is a j in $i+1, \ldots, d-1$ with $(m_{k,i}, \{m_{k+1,i}, \ldots, m_{d,i}\}) \in B_M(j)$ or $\in \tilde{B}_M(j)$, where $B_M(j) := \{(m_{j,j}, D) | k = j+1, \ldots, d\}$ with $D = \{m_{k,j}, \ldots, m_{d,j}\}$ and $\tilde{B}_M(j) := \{(m_{k,j}, \tilde{D}) | k = j+1, \ldots, d\}$ with $\tilde{D} = \{m_{j,j}\} \cup \{m_{k+1,j}, \ldots, m_{d,j}\}$.*

For the R-Vine copula example in Fig. 1.1, the R-Vine matrix $\mathbf{M}^*$ is given as

$$\begin{bmatrix} 5 & & & & \\ 4 & 4 & & & \\ 1 & 1 & 1 & & \\ 2 & 3 & 3 & 3 & \\ 3 & 2 & 2 & 2 & 2 \end{bmatrix},$$

where the first column represents the dependence of four pairs of variables, $(5, 4|123)$, $(5, 1|23)$, $(5, 2|3)$ and $(5, 3)$. Going through all columns, we can see that the matrix $\mathbf{M}^*$ codes all information needed to represent the R-vine copula in Fig. 1.1.

An R-Vine array has the following two properties:

- $\{m_{i,i}, \ldots, m_{d,i}\} \subset \{m_{j,j}, \ldots, m_{d,i}\}$ for $1 \leq j < i \leq d$,
- $m_{i,i} \notin \{m_{i+1,i+1}, \ldots, m_{d,i+1}\}$ for $i = 1, \ldots, d-1$,

where the first property states that every column in the left contains all the entries that a column in the right contains, and the second property guarantees that there is a new entry on the diagonal in every column.

Given an R-Vine array $M = (m_{i,j})_{i,j=1,\ldots,d}$, the R-Vine copula based modeling of the joint PDF [27] is

$$f_{1,\ldots,d} = \prod_{j=1}^{d} f_j \prod_{k=d-1}^{1} \prod_{i=d}^{k+1} c_{m_{k,k},m_{i,k}|m_{i+1,k},\ldots,m_{d,k}} \left(F_{m_{k,k}|m_{i+1,k},\ldots,m_{d,k}}, F_{m_{i,k}|m_{i+1,k},\ldots,m_{d,k}} \right). \tag{1.14}$$

For notational simplicity, we have removed the arguments of all the functions in Equation (1.14).

1.2 Literature Review

Multimodal signal processing enables fusion of information from several sources in order to form a unified picture and produce a global decision/estimation. There are mainly three fusion strategies: data-level fusion, feature-level fusion and decision-level fusion. Signal processing for inference problems with distributed sensors has been studied extensively. *Centralized* inference (also known as data-level fusion), where raw observations are available at the processing unit or FC, have been well studied in standard textbooks [13, 59, 101]. *Distributed* inference, on the other hand, relies on the topology of a network that can either transmit a compressed/processed version of the raw data to the FC (can be feature-level fusion or decision-level fusion) or arrive at a consensus solution by locally sharing compressed/processed information (e.g., see [55, 56, 67, 78, 105, 117] and references cited therein).

This section reviews some recent progress that has been made in the field of multimodal signal processing, and focuses on developments where data dependence plays a significant role in the design of fusion rules for inference problems. The aim of this discussion is to motivate our current research.

1.2.1 Linear Dependence: Covariance Matrix

Covariance matrix, or equivalently correlation matrix, models linear dependence among jointly normal random variables or variables that possess a finite second moment. In networks with multiple sensors/sources, it is used extensively to model dependency information across the sensors/sources, especially where it is reasonable to assume linearity of the medium of signal propagation. In MIMO systems, the dependence among multiple antennas/channels was modeled in [52, 63, 71]. In [4], linear dependence among multiple datasets was characterized for joint blind source separation. Canonical correlation analysis (CCA) has also been used to perform feature-level information fusion for recognition problems [29, 35, 39].

Optimal schemes for distributed detection and estimation with dependent observations have also been a topic of significant interest. In the case of distributed detection, it has been shown in [105] that the optimal sensor decision rule is the likelihood-ratio-based binary quantizer, and the optimal fusion statistic at the FC is a weighted sum of sensor decisions under the assumption of conditional independence. These sensor decision rules and fusion statistic are no longer optimal when correlation is taken into account. Examples of the consequent loss in performance were presented in [1]. Moreover, it has been shown in [102] that the distributed detection problem with dependent observations is NP-complete and cannot be solved using a polynomial time algorithm. Therefore, the design of optimal local decision rules may not be possible due to computational intractability resulting from the dependence among sensor observations. One way to get past the computational intractability is to assume some prior information about the joint statistics, e.g., in [28, 57], fusion rules for correlated binary decisions were studied by considering known correlation coefficients and known correlated sensing noise PDFs, respectively. Another way is to constrain local detectors to be binary quantizers and design optimal fusion rules at the FC, e.g., in [18, 109], optimal fusion rules were proposed with correlated Gaussian noise. Also, in [58], noisy correlated sensing channels were studied for multi-bit decision based distributed detection and a likelihood ratio test was used to generate

the global decision at the FC.

The distributed estimation problem by modeling dependent observations has been studied in [32, 61, 66]. In [61], a distributed estimation scheme was studied with multivariate Gaussian correlated sensor observations and the covariance was assumed to be known at the FC. In [32], the estimation of a random scalar parameter in a power constrained wireless sensor network was studied with generally correlated sensor observations that can accommodate nonlinear measurement models and spatially correlated observation noise. The goal was to design optimal power allocation strategy. In [66], the problem of sensor selection for parameter estimation was considered with spatially correlated Gaussian measurement noise and the aim was to seek optimal sensor activations by formulating an optimization problem which minimizes estimation error subject to energy constraints. Besides these formulations, designing estimation schemes in the presence of dependent data often gives rise to intractable problem formulations. In such situations, applying well-known strategies derived from conditional independence assumption may turn out to be fairly suboptimal. One way to address this issue is to allow inter-sensor communication/collaboration instead of modeling this dependence [16, 20, 31, 55, 56, 67, 87]. In [31, 56, 67], collaborative distributed estimation problems with a fusion center were considered, where collaboration was restricted to be a linear operation. Collaborative distributed estimation problems without a FC were studied in [16, 20, 55, 87], where different distributed collaboration strategies were proposed, such as diffusion-based, consensus-based and gossip-based algorithms.

1.2.2 Nonlinear Dependence: Nonparametric Approach

Multimodal signal processing using nonparametric approaches has attracted significant attention in applications where it is infeasible to model the complex nonlinear dependencies that may exist among sensor observations/features. These methods, in essence, estimate or learn the joint distribution across sensor observations/features directly from the data.

Information theoretic approaches make it possible to characterize arbitrary nonlinear dependence compared to methods using covariance matrix. In [15], mutual information and joint entropy based methodologies were proposed to model the underlying dependence between audio and video data. In [14,40,85], mutual information based methods were proposed for image fusion. Graphical models such as Bayesian networks generalize hidden Markov models and have also been successfully used for multimodal fusion (see e.g. [23,53,81,96]).

Machine learning and deep learning techniques have had breakthroughs in a wide range of multimodal applications: from audio-visual speech recognition to image captioning [7, 8, 38, 86]. The advantage of machine learning and deep learning based methodologies is that they can extract significant amount of information from sensor data with no need of modeling the joint distribution of the data. There are plenty of networks including shallow networks, such as Support Vector Machines (SVMs), Random Forests and Decision Trees, and deep networks, such as deep forward neural networks and convolutional neural networks. Compared to the shallow networks, the deep networks can learn high-level features directly from raw data (or lightly processed data) and provide joint representations for multimodal data.

1.2.3 Nonlinear Dependence: Copula-based Approach

Recall from Section 1.1 that copulas are parametric functions that can model nonlinear dependence among multiple random variables. The copula based dependence modeling approach is attractive and powerful because it can characterize potentially any nonlinear spatial dependence among sensor observations and allow different marginal distributions. Moreover, while nonparametric approaches have shown their superiority in characterizing the joint distribution among multimodal data, they also suffer from issues, such as scalability problems stemming from the curse of dimensionality (information theoretic/graphical model based approaches) and the availability of enough training data (deep learning based approaches). Recently, considerable progress has been made in the study of copulas and their applications in statistics.

The usage of copulas is widespread in the fields of econometrics and finance [19] and they are beginning to be used in the signal and image processing context [24, 42, 48, 70, 93].

Multivariate copula based approaches have shown their superiority in improving the performance of inference problems [43, 45, 50, 94, 95]. In [50], a general framework of copula based detection has been investigated. The performance loss due to copula misspecification was quantified. The efficacy of the proposed copula based detection scheme was demonstrated using a NIST multibiometric dataset. In [95], the problem of distributed detection has been studied, where a copula based optimum fusion rule was derived for a Neyman-Pearson detector. In [45], the utility of non-stationary dependence modeling with copulas has been considered for detecting the presence of a phenomenon being observed jointly by heterogeneous sensors. In [94], a copula-based estimation scheme has been proposed for the localization of a radiation source, and the overall estimation performance was shown to be improved by taking the underlying dependence among sensor observations into account. In [43], the fusion of social media and sensor data has been addressed using the copula-based dependence modeling approach.

However, the class of known multivariate copulas required for the fusion of observations from more than two sensors is limited. Gaussian copulas perform poorly on data with heavy tails. Student-t copulas allow for symmetric tail dependence, but they have only a single parameter to capture tail dependence among all the variables. While standard Archimedean multivariate copulas can characterize asymmetric tail dependence, they are quite limited as they are characterized by only a single parameter. This shows that there is a growing need for more flexible copulas especially for modeling high-dimensional dependence structures. Vine copulas, tree-structured graphical models, are more flexible and powerful compared to multivariate copulas, where a set of bivariate copulas are used to construct the multivariate copula [2, 11, 12].

1.3 Main Contributions and Organization

The main contributions of the research results presented in this dissertation to the signal processing and information fusion literature, are as follows:

In Chapter 2, we propose a regular vine copula based methodology for the fusion of statistically dependent decisions. Regular vine copula can express a multivariate copula by using a cascade of bivariate copulas, the so-called pair copulas. Assuming that local detectors are single threshold binary quantizers and taking complex dependence among sensor decisions into account, we design an optimal fusion rule using a regular vine copula under the Neyman-Pearson framework. In order to reduce the computational complexity resulting from the complex dependence, we propose an efficient and computationally light regular vine copula based optimal fusion algorithm. Numerical experiments are conducted to demonstrate the effectiveness of our approach.

Nowadays, we are inundated by a large amount of streaming data that are generated continuously with high arrival rates from sources such as sensors and social media. The methods applied for processing data streams should be fast enough to keep up with the high arrival rate of incoming data, and at the same time provide solutions for inference problems (detection, classification, or estimation) with high accuracies. In Chapter 3, we design a novel parallel platform, C-Storm (Copula-based Storm), for the computationally complex problem of fusion of heterogeneous data streams for inference. C-Storm is designed by marrying copula-based dependence modeling for highly accurate inference and a highly-regarded parallel computing platform Storm for fast stream data processing. C-Storm has the following desirable features: 1) C-Storm offers fast inference responses. 2) C-Storm provides high inference accuracies. 3) C-Storm is a general-purpose inference platform that can support data fusion applications. 4) C-Storm is easy to use and its users do not need to have deep knowledge of Storm or copula theory.

In Chapter 4, we study the problem of multi-sensor based human activity recognition via

the fusion of multiple deep neural network classifiers. We take the cross-modal dependence into account by employing regular vine copulas to characterize complex dependence among multiple modalities. More specifically, multiple deep neural networks are used to extract high-level features from multiple sensing modalities, with each deep neural network processing the data collected from a single sensor. The extracted high-level features are then combined using a regular vine copula model. Numerical experiments are conducted to demonstrate the effectiveness of our approach.

In Chapter 5, we consider the problem of distributed sequential detection using wireless sensor networks in the presence of imperfect communication channels between the sensors and the fusion center. Sensor observations are assumed to be spatially dependent. Moreover, the channel noise can be dependent and non-Gaussian. We propose a copula based distributed sequential detection scheme that takes the spatial dependence into account. More specifically, each local sensor runs a memory-less truncated sequential test repeatedly and sends its binary decisions to the fusion center synchronously. The fusion center fuses the received messages using a copula-based sequential test. To this end, we first propose a centralized copula based sequential test and show its asymptotic optimality and time efficiency. We then show the asymptotic optimality and time efficiency of the proposed distributed scheme. We also show that by suitably designing the local thresholds and the truncation window, the local probabilities of false alarm and miss detection of the proposed memory-less truncated local sequential tests satisfy the pre-specified error probabilities. Numerical experiments are conducted to demonstrate the effectiveness of our approach.

In Chapter 6, we consider the problem of collaborative distributed estimation in a large scale sensor network with statistically dependent sensor observations. In the collaborative setup, the aim is to maximize the overall estimation performance by modeling the underlying statistical dependence and efficiently utilizing the deployed sensors. To achieve greater sensor transmission and estimation efficiency, we propose a two-step cluster-based collaborative

distributed estimation scheme, where in the first step, sensors form dependence driven clusters such that sensors in the same cluster are dependent, while sensors from different clusters are independent, and perform copula-based maximum a posteriori probability (MAP) estimation via intra-cluster collaboration. In the second step, the estimates generated in the first step are shared via inter-cluster collaboration to reach an average consensus. A merge based K-medoid dependence driven clustering algorithm is proposed. Moreover, we further propose a cluster-based sensor selection scheme using mutual information prior to the estimation. The aim is to select sensors with maximum relevance and minimum redundancy regarding the parameter of interest under certain pre-specified energy constraint. Also, the proposed cluster-based sensor selection scheme is shown to be equivalent to the global/non-cluster based selection scheme with high probability, which at the same time is computationally more efficient. Numerical simulations are conducted to demonstrate the effectiveness of our approach.

Finally, in Chapter 7, we summarize the findings and results of this dissertation. Several directions and ideas for future research are also presented.

1.4 Bibliographic Note

Part of the work presented in this dissertation has appeared in the following publications:

1. Shan Zhang, Jielong Xu, Sora Choi, Jian Tang, Pramod K. Varshney, and Zhenhua Chen, “A Parallel Platform for Fusion of Heterogeneous Stream Data," in *Proc. 21th International Conference on Information Fusion*, 2018.

2. Shan Zhang, Baocheng Geng, Pramod K. Varshney, and Muralidhar Rangaswamy, “Fusion of Deep Neural Networks for Activity Recognition: A Regular Vine Copula Based Approach," in *Proc. 22th International Conference on Information Fusion*, 2019.

3. Shan Zhang, Prashant Khanduri, and Pramod K. Varshney, “Distributed Sequential Hypothesis Testing with Dependent Sensor Observations," in *Proc. 53rd IEEE Asilomar*

Conference on Signals Systems and Computer, 2019.

4. Shan Zhang, Lakshmi N. Theagarajan, Sora Choi, and Pramod K. Varshney, “Fusion of Correlated Decisions Using Regular Vine Copulas," *IEEE Transactions on Signal Processing*, vol 67, no. 8, pp. 2066-2079, 2019.

5. Shan Zhang, Prashant Khanduri, and Pramod K. Varshney, “Distributed Sequential Detection: Dependent Observations and Imperfect Communication," *IEEE Transactions on Signal Processing*, accepted, 2019.

6. Shan Zhang, Pranay Sharma, and Pramod K. Varshney, “Distributed Estimation in Large Scale Wireless Sensor Networks via A Two-Step Cluster-based Approach," to be submitted, 2019.

CHAPTER 2
DISTRIBUTED DETECTION BASED ON REGULAR VINE COPULAS

2.1 Motivation

Fusion of data from heterogeneous sensors/sources has been shown to improve the performance of inference tasks. In many practical cases, these sensor observations are dependent due to a variety of reasons such as sensing of the same phenomenon and dependent transmission channels. Ignoring this dependence may degrade inference performance.

In this chapter, we consider the problem of distributed detection with dependent sensor observations under the Neyman-Pearson framework. We assume that local detectors are single threshold binary quantizers, and the aim is to derive an optimal fusion rule at the FC by taking the dependent decisions into consideration. We propose a novel and powerful fusion methodology for the fusion of dependent decisions, R-Vine copula based fusion, for more flexible modeling of complex dependency especially for larger number of sensors. In order to reduce the computational complexity resulting from the complex dependence, we further propose an efficient and computationally light regular vine copula based optimal fusion algorithm.

2.2 Problem Formulation

Consider a distributed detection problem, where a random phenomenon is monitored by L sensors. A binary hypothesis testing problem is studied, where H_1 denotes the presence of the random phenomenon and H_0 denotes the absence of the phenomenon. The sensors make a set of observations at time instant n, $\mathbf{z}_n = [z_{1n}, z_{2n}, \ldots, z_{Ln}], n = 1, 2, \ldots, N$. We assume that the sensor observations are dependent across sensors. Moreover, we further assume that the sensor observations are continuous random variables that are conditionally independent and identically distributed (i.i.d.) over time. Let $f(z_{ln}|H_1)$ and $f(z_{ln}|H_0)$ be the PDFs of the observation at the lth sensor and nth time instant under H_1 and H_0 hypotheses, respectively. No knowledge about the joint distribution of the sensor observations is available *a priori*. Instead of transmitting noisy raw observations, local binary sensor decisions u_{ln} are sent to the FC by using a binary quantizer which is defined as

$$u_{ln} = \begin{cases} 0 & -\infty < z_{ln} < \tau_l \\ 1 & \tau_l \leq z_{ln} < +\infty \end{cases}, \tag{2.1}$$

where τ_l is the quantizer threshold at the lth sensor. At the FC, local binary decisions are combined to obtain a global decision.

Under the Neyman-Pearson criterion, the design problem for the parallel distributed detection system consists of deriving individual sensor thresholds τ_l to form sensor decisions and the optimal fusion rule that fuses local sensor decisions to obtain the global decision. The sensor thresholds $\tau_l, l = 1, 2, \ldots, L$ are obtained by maximizing the local probability of detection subject to a constraint on the local probability of false alarm. Note that these sensor thresholds are not necessarily optimal in the global sense. The design of the optimal fusion rule for multiple sensors is discussed next.

Since sensor decisions are independent over time, the optimal test statistic [104] is given as

$$\Lambda(\mathbf{u}) = \frac{\prod_{n=1}^{N} P(u_{1n}, u_{2n}, \ldots, u_{Ln}|H_1)}{\prod_{n=1}^{N} P(u_{1n}, u_{2n}, \ldots, u_{Ln}|H_0)}, \tag{2.2}$$

where $P(u_{1n}, u_{2n}, \ldots, u_{Ln}|H_k)$ is the joint probability mass function (PMF) of the sensor decisions at the nth time instant under kth hypothesis, $k = 0, 1$. We define $S = \{u_{1n}u_{2n}\ldots u_{Ln}|u_{ln} \in \{0, 1\}, l = 1, 2, \ldots, L\}$ as the set of all permutations that specify L-sensor decisions at time instant n. There are a total of 2^L permutations for L sensors. For a three-sensor problem, $S = \{\{000\}, \{001\}, \{010\}, \{011\}, \{100\}, \{101\}, \{110\}, \{111\}\}$. Let

$$\begin{aligned} P(u_{1n}, u_{2n}, \ldots, u_{Ln}|H_1) &= P_s, \text{ and} \\ P(u_{1n}, u_{2n}, \ldots, u_{Ln}|H_0) &= Q_s, \end{aligned} \tag{2.3}$$

where $s \in S$. P_s and $Q_s, s \in S$ are required while computing the test statistic at the FC. For a three-sensor problem, the set of probabilities P_{000}, P_{001}, P_{010}, ..., P_{111} and Q_{000}, Q_{001}, Q_{010}, ..., Q_{111} that characterize the joint PMFs of sensor decisions u_{1n}, u_{2n} and u_{3n} under hypotheses H_1 and H_0, respectively, are needed. By integrating the joint PDFs of the sensor observations under both hypotheses, these probabilities can be obtained with the quantizer threshold τ_l, $l = 1, 2, 3$. For example,

$$\begin{aligned} P_{000} &= \int_{z_1=-\infty}^{\tau_1} \int_{z_2=-\infty}^{\tau_2} \int_{z_3=-\infty}^{\tau_3} f(z_1, z_2, z_3|H_1) dz_1 dz_2 dz_3, \\ P_{010} &= \int_{z_1=-\infty}^{\tau_1} \int_{\tau_2}^{z_2=+\infty} \int_{z_3=-\infty}^{\tau_3} f(z_1, z_2, z_3|H_1) dz_1 dz_2 dz_3, \end{aligned} \tag{2.4}$$

where for the simplification of notation, we omit the time index n in the example.

However, due to existing complex and nonlinear dependence, the joint PDFs of sensor observations under both hypotheses are not known. Before determining the joint PMFs of sensor decisions, we first need to obtain the joint PDFs of sensor observations given only the knowledge of marginal PDFs of the sensor observations and the marginal PMFs of sensor

decisions. Typically in many applications, we do not have any prior information related to the phenomenon of interest. Therefore, we may also need to determine the marginals of sensor observations.

The dependence across sensors can be quite complicated and nonlinear. Simple dependence modeling through methods such as the use of multivariate normal model, is very limited and inadequate to characterize complex dependence among multiple sensors. Assuming conditional independence among multiple sensors may result in substantial performance degradation. To design the optimal fusion rule, we propose a copula based fusion methodology to characterize the existing dependence and determine the joint PDFs of sensor observations. Due to the limitations of the class of standard multivariate copulas and complex dependence that generally exists among multiple sensors, more flexible dependence modeling approaches are needed to obtain the joint PDFs of sensor measurements. R-Vine copula based dependence modeling provides us a solution. It can express a multivariate copula using a cascade of bivariate copulas embedded in a tree structure that is shown to be more flexible and powerful to model the complex dependence. Note that learning of the joint distribution requires raw sensor observations. It can be done offline. Here, we assume that the joint statistics of the sensors does not change over time. After measurement collection, raw measurements are sent to the FC. The FC uses these analog measurements to learn the joint statistics of the sensors. After that, only binary decisions are sent to the FC.

Taking the above considerations into account, in the following, we develop a novel and powerful R-Vine copula based fusion methodology for distributed detection. We will propose the optimal test statistic for the parallel distributed detection system and derive its asymptotic statistic when the number of observations is large. Furthermore, at the end, via simulations, we will show its power and flexibility to capture complex dependence and improve detection performance.

2.3 R-Vine Copula Based Fusion of Multiple Statistically Dependent Decisions

2.3.1 Optimal Test Statistic

The optimal test statistic for L sensors is characterized in Equation (2.2). The joint PMF of u_{ln}, $l = 1, 2, \ldots, L$, at time n, $n = 1, 2, \ldots, N$ under H_1 and H_0, respectively, is given as:

$$\begin{aligned} P(u_{1n}, u_{2n}, \ldots, u_{Ln}|H_1) &= \prod_{s\in S} P_s^{\prod_{l=1}^{L} x_{ln}}, \\ P(u_{1n}, u_{2n}, \ldots, u_{Ln}|H_0) &= \prod_{s\in S} Q_s^{\prod_{l=1}^{L} x_{ln}}, \end{aligned} \tag{2.5}$$

where s_l indicates the lth element of s, and $x_{ln} = u_{ln}$ if $s_l = 1$, otherwise, $x_{ln} = 1 - u_{ln}$ for $s \in S$. For example, please see Equation (2.7) and Equation (2.8), which are special cases of Equation (2.5) for $L = 3$.

Substituting Equation (2.5) in Equation (2.2) and taking log on both sides, the log test statistic is given by

$$\begin{aligned} \log\Lambda(\mathbf{u}) = \sum_{\{i_1 n\}\in I_1} A_{\mathbf{u}^1} \sum_{n=1}^{N} \mathbf{u}^1 \; + \sum_{\{i_1 n, i_2 n\}\in I_2} A_{\mathbf{u}^2} \sum_{n=1}^{N} \mathbf{u}^2 + \ldots + \\ \sum_{\{i_1 n, i_2 n, \ldots, i_t n\}\in I_t} A_{\mathbf{u}^t} \sum_{n=1}^{N} \mathbf{u}^t + \ldots + \sum_{\{i_1 n, i_2 n, \ldots, i_L n\}\in I_L} A_{\mathbf{u}^L} \sum_{n=1}^{N} \mathbf{u}^L \end{aligned} \tag{2.6}$$

where $I = \{ln | u_{ln} \in \{0, 1\}, l = 1, 2, \ldots, L, n = 1, 2, \ldots, N\}$, I_i is a subset of I and the cardinality of the set I_i is i, namely, $|I_i| = i$. Moreover, $\mathbf{u}^t = \{u_{i_1 n} u_{i_2 n} \ldots u_{i_t n}\}$, $t \in [1, 2, \ldots, L]$ and its weight is given as $A_{\mathbf{u}^t} = \log \frac{\prod_{0\le k\le t} \mathcal{P}_{\tilde{I}^e_{tk}}^{(-1)^t} \prod_{0\le k\le t} \mathcal{Q}_{\tilde{I}^o_{tk}}^{(-1)^t}}{\prod_{0\le k\le t} \mathcal{Q}_{\tilde{I}^e_{tk}}^{(-1)^t} \prod_{0\le k\le t} \mathcal{P}_{\tilde{I}^o_{tk}}^{(-1)^t}}$ which is determined by the joint PMFs of sensor decisions, see Appendix A for details. Also, please see Equation (2.9) as an example for $L = 3$.

The optimal test statistic for the three-sensor case

Considering the three-sensor case, the joint PMF of u_{1n}, u_{2n} and u_{3n} at any time instant, $1 \leq n \leq N$, under H_1 and H_0 is given as follows, respectively,

$$\begin{aligned} &P(u_{1n}, u_{2n}, u_{3n}|H_1) = \\ &P_{000}^{(1-u_{1n})(1-u_{2n})(1-u_{3n})} P_{001}^{(1-u_{1n})(1-u_{2n})u_{3n}} P_{010}^{(1-u_{1n})u_{2n}(1-u_{3n})} \\ &P_{011}^{(1-u_{1n})u_{2n}u_{3n}} P_{100}^{u_{1n}(1-u_{2n})(1-u_{3n})} P_{101}^{u_{1n}(1-u_{2n})u_{3n}} \\ &P_{110}^{u_{1n}u_{2n}(1-u_{3n})} P_{111}^{u_{1n}u_{2n}u_{3n}}, \end{aligned} \tag{2.7}$$

and

$$\begin{aligned} &P(u_{1n}, u_{2n}, u_{3n}|H_0) = \\ &Q_{000}^{(1-u_{1n})(1-u_{2n})(1-u_{3n})} Q_{001}^{(1-u_{1n})(1-u_{2n})u_{3n}} Q_{010}^{(1-u_{1n})u_{2n}(1-u_{3n})} \\ &Q_{011}^{(1-u_{1n})u_{2n}u_{3n}} Q_{100}^{u_{1n}(1-u_{2n})(1-u_{3n})} Q_{101}^{u_{1n}(1-u_{2n})u_{3n}} \\ &Q_{110}^{u_{1n}u_{2n}(1-u_{3n})} Q_{111}^{u_{1n}u_{2n}u_{3n}}. \end{aligned} \tag{2.8}$$

For simplification of notation, we use A_1 to A_7 to denote the coefficients of $\mathbf{u}^t$, $t = 1, 2, 3$. Substituting Equation (2.7) and Equation (2.8) into Equation (2.2) and taking log on both sides, we get

$$\begin{aligned} &\log\Lambda_1(\mathbf{u}) = \\ &A_1 \sum_{n=1}^{N} u_{1n} + A_2 \sum_{n=1}^{N} u_{2n} + A_3 \sum_{n=1}^{N} u_{3n} + A_4 \sum_{n=1}^{N} u_{1n}u_{2n} + \\ &A_5 \sum_{n=1}^{N} u_{1n}u_{3n} + A_6 \sum_{n=1}^{N} u_{2n}u_{3n} + A_7 \sum_{n=1}^{N} u_{1n}u_{2n}u_{3n}, \end{aligned} \tag{2.9}$$

where

$$A_1 = \log\frac{Q_{000}P_{100}}{P_{000}Q_{100}}, \qquad A_2 = \log\frac{Q_{000}P_{010}}{P_{000}Q_{010}},$$
$$A_3 = \log\frac{Q_{000}P_{001}}{P_{000}Q_{001}}, \qquad A_4 = \log\frac{P_{000}Q_{100}Q_{010}P_{110}}{Q_{000}P_{100}P_{010}Q_{110}},$$
$$A_5 = \log\frac{P_{000}Q_{100}Q_{001}P_{101}}{Q_{000}P_{100}P_{001}Q_{101}}, A_6 = \log\frac{P_{000}Q_{010}Q_{001}P_{011}}{Q_{000}P_{010}P_{001}Q_{011}},$$
$$A_7 = \log\frac{Q_{000}P_{100}P_{010}P_{001}Q_{110}Q_{101}Q_{011}P_{111}}{P_{000}Q_{100}Q_{010}Q_{001}P_{110}P_{101}P_{011}Q_{111}}.$$

When sensor decisions among L sensors are conditionally independent, only the term $\sum_{\{i_1 n\}\in I_1} A_{\mathbf{u}^1} \sum_{n=1}^{N} \mathbf{u}^1$ in Equation (2.6) is left and the optimal fusion rule reduces to the Chair-Varshney fusion rule statistic (i.e., weighted sum of sensor decisions [17]). For dependent sensor decisions, the optimal fusion rule depends on both the weighted sum of sensor decisions and the weighted sum of the cross products of sensor decisions. The cross products of the sensor decisions are due to dependence among multiple sensors. The joint PMFs of sensor decisions, namely P_s and Q_s, $s \in S$, determine the weights of the optimal test statistic, and can be obtained by solving L integrals on the joint PDFs of the corresponding sensor observations (see the example in Equation (2.4)). In the following subsection, we will propose an R-Vine copula based approach to model existing complex dependence and construct the joint PDFs of sensor observations. After obtaining the joint PMFs and given sensor decisions, the optimal fusion rule is given by

$$\log\Lambda(\mathbf{u}) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma, \tag{2.10}$$

where γ is the threshold for the test at the FC.

To characterize the fusion performance at the FC using the system probabilities of detection and false alarm, we consider the asymptotic distribution of the optimal fusion rule statistic under H_0 and H_1.

Theorem 2.1. *The optimal fusion test statistic $\log\Lambda(\mathbf{u})$ is asymptotically (when N is large) Gaussian.*

Proof: See Appendix B. ■

The first and second order statistics of $\log\Lambda(\mathbf{u})$ under both hypotheses are given in Appendix B. Let the first and second order statistics of $\log\Lambda(\mathbf{u})$ be denoted by μ_0 and σ_0^2 under H_0 and μ_1 and σ_1^2 under H_1. These can be easily derived using the joint PMFs of sensor decisions. The system probability of detection (P_D) and system probability of false alarm (P_F) are then given by

$$P_D = Q\left(\frac{\gamma - \mu_1}{\sigma_1}\right), \tag{2.11}$$

$$P_F = Q\left(\frac{\gamma - \mu_0}{\sigma_0}\right), \tag{2.12}$$

where $Q(\cdot)$ is the complementary CDF of the Gaussian distribution. Under the Neyman-Pearson framework and by constraining $P_F = \alpha$, γ can be obtained by

$$\gamma = \sigma_0 Q^{-1}(P_F) + \mu_0. \tag{2.13}$$

Note that the local sensors compress their raw measurements into binary decisions (see Equation (2.1)) prior to their transmission to the FC and the corresponding sensor thresholds are assumed to be $\tau_l, l = 1, 2, \ldots, L$. Let $\boldsymbol{\tau}$ be the vector of sensor thresholds. Constraining $P_F = \alpha$, P_D can be written as

$$P_D(\boldsymbol{\tau}) = Q\left(\frac{\sigma_0 Q^{-1}(P_F) + \mu_0(\boldsymbol{\tau}) - \mu_1(\boldsymbol{\tau})}{\sigma_1(\boldsymbol{\tau})}\right), \tag{2.14}$$

where $\boldsymbol{\tau}$ is chosen to maximize P_D at a particular value of P_F.

It should be noted that the computational complexity for obtaining the joint PMFs is very high since we need to perform multi-dimensional integration at each time instant. In what follows, we first propose the R-Vine copula based methodology to characterize the joint PDFs of sensor observations and then develop an efficient optimal fusion algorithm based on the R-Vine copula model.

2.3.2 R-Vine Copula Based Dependence Modeling

According to Sklar's theorem (Section 1.1.1), the joint PDF of sensor observations can be separated into its marginals and the dependence structure that is fully characterized by the copula density (see Equation (1.2)). As indicated earlier, the R-Vine copula model (Section 1.1.4) is more flexible to decompose the joint PDF into its marginals and a cascade of bivariate copula densities. In the following, we will use the R-Vine copula to model the dependence structure and obtain the joint PDF of sensor observations.

In our parallel distributed detection sensor network, L sensors make a set of observations $\mathbf{z}_n = [z_{1n}, \ldots, z_{Ln}]$ at time instant n. Recall that we assume the sensor observations to be conditionally i.i.d. over time. Therefore, it is sufficient to consider the joint PDF of $\mathbf{z}_n$. For notational convenience, we omit the index n in this subsection and let $\mathbf{z} = [z_1, \ldots, z_L]$ be the L-dimensional observation vector with its marginal CDFs, $\mathbf{F} = [F_1(z_1), \ldots, F_L(z_L)]$. The R-Vine copula $(\mathbf{F}, \mathcal{V}, \mathbf{B})$ (see Definition 1.2) of $\mathbf{z}$ is specified by its marginal CDFs $\mathbf{F}$, R-Vine $\mathcal{V} = (T_1, \ldots, T_{L-1})$ and a set of bivariate copulas $\mathbf{B} = \{C_{\mathfrak{e}_{e,a},\mathfrak{e}_{e,b}|D_e} \,|\, e \in E_i, i = 1, 2, \ldots, L-1\}$ with a set of parameters $\boldsymbol{\phi}$.

From Equation (1.11), the joint PDF of $\mathbf{z}$ is given as

$$f(\mathbf{z}|\mathcal{V}, \mathbf{B}, \boldsymbol{\phi}) = \prod_{l=1}^{L} f(z_l) \prod_{i=1}^{L-1} \prod_{e \in E_i} \times \tag{2.15}$$

$$c_{\mathfrak{e}_{e,a},\mathfrak{e}_{e,b}|D_e}(F_{\mathfrak{e}_{e,a}|D_e}(z_{\mathfrak{e}_{e,a}}|\mathbf{z}_{D_e}), F_{\mathfrak{e}_{e,b}|D_e}(z_{\mathfrak{e}_{e,b}}|\mathbf{z}_{D_e}); \boldsymbol{\phi}),$$

where $e = \{a, b\}$, $\mathbf{z}_{D_e} = \{z_j | j \in D_e\}$, $f(z_l)$ is the marginal PDF of the observation of sensor l, $l = 1, \ldots, L$.

Given a set of N observed data $\mathbf{z}_1, \ldots, \mathbf{z}_N$, the joint PDF of the observations is given as

$$f(\mathbf{z}_1, \ldots, \mathbf{z}_N) = \prod_{n=1}^{N} f(\mathbf{z}_n|\mathcal{V}, \mathbf{B}, \boldsymbol{\phi}). \tag{2.16}$$

2.3.3 Model Selection and Estimation

The fitting of an R-Vine copula model to given data requires the selection of the R-Vine tree structure $\mathcal{V}$, the choice of copula families for the bivariate copula set $\mathbf{B}$ and the estimation of their corresponding parameters $\boldsymbol{\phi}$. Since the bivariate copula families and their corresponding parameters both depend on the R-Vine tree structure, the identification of trees accurately is key to the R-Vine copula model. It has been shown that the number of possible R-Vines for n variables increases very rapidly and is given by $\binom{n}{2} \times (n-2)! \times 2^{\binom{n-2}{2}}$ [73]. It is not computationally feasible to find the best model by fitting all possible R-Vine constructions. Suboptimal R-Vine copula selection strategies have been investigated in the literature. In [27], a sequential method to select an R-Vine model based on Kendall's tau was proposed, where a maximum spanning tree algorithm was used. Moreover, the feasibility and efficiency of this method was demonstrated. The sequential method starts with the selection of the first tree T_1 and continues tree by tree up to the last tree T_{L-1}. The trees are selected in a way that the chosen bivariate copula models the strongest pair-wise dependencies present which are characterized by Kendall's tau. There are other possible choices to measure the pair-wise dependencies besides Kendall's tau, for example, the Akaike Information Criterion (AIC) [3] of each bivariate copula proposed in [21] and the p-value of a copula goodness of fit test and variants proposed in [22].

In this chapter, we adopt the sequential method proposed in [27] to construct the R-Vine copula model. Also, we use Kendall's tau as the measure of dependencies and select the spanning tree that maximizes the sum of the absolute values of empirical Kendall's tau. Kendall's tau can be expressed as an expectation over a bivariate copula distribution as shown in [75], and typically, the log likelihood of a bivariate copula increases with increasing absolute values of Kendall's tau. Moreover, the advantage of using Kendall's tau is that one does not need to select and estimate the bivariate copulas prior to the tree selection step. We summarize the sequential method based on Kendall's tau for obtaining the joint PDF of sensor observations in

Algorithm 2.1.

Besides the selection of the R-Vine tree structure, we need to define a copula family for each pair of sensors and select the copula that best characterizes the pair-wise dependencies. Consider a library of copulas, $\mathcal{C} = \{c_m : m = 1, \ldots, M\}$ and assume that we have a set of N observations $\mathbf{z}_1, \ldots, \mathbf{z}_N$. Based on Equation (2.15), to obtain the joint PDF of sensor observations, we need to specify the marginal PDFs, marginal CDFs including conditional marginal CDFs of individual local sensor observations as well as the bivariate dependence structure. If we do not have any prior knowledge of the phenomenon of interest, the marginal PDFs $f(z_{ln})$ for sensor $l, l = 1, 2, \ldots, L$ at time instant $n, n = 1, 2, \ldots, N$ can be estimated non-parametrically using Kernel density estimators [108], and the marginal CDFs $F(z_{ln})$ can be determined by the Empirical Probability Integral Transforms (EPIT) [45]. Note that the conditional marginal CDFs need to be obtained recursively using Equation (1.12). Before selecting the best bivariate copula, the copula parameter set $\boldsymbol{\phi}$ is obtained using MLE, which is given by

$$\widehat{\boldsymbol{\phi}} = \arg\max_{\boldsymbol{\phi}} \sum_{n=1}^{N} \log c(F(z_{l_1 n}), F(z_{l_2 n})|\boldsymbol{\phi}), \tag{2.17}$$

where $(l_1, l_2), l_1, l_2 \in [1, 2, \ldots, L]$ is a connected pair in R-Vine tree $\mathcal{V}$ and for simplification of notation, we omit the conditioned elements for conditional marginal CDFs.

To decide on the best copula, we consider three widely used model selection criteria: AIC, Bayesian Information Criterion (BIC) [88], and MLE,

$$\begin{aligned}
\text{AIC} &= -\sum_{n=1}^{N} \log c(F(z_{l_1 n}), F(z_{l_2 n})|\widehat{\boldsymbol{\phi}}) + 2q_c, \\
\text{BIC} &= -\sum_{n=1}^{N} \log c(F(z_{l_1 n}), F(z_{l_2 n})|\widehat{\boldsymbol{\phi}}) + q_c \log(N), \\
\text{MLE} &= \sum_{n=1}^{N} \log c(F(z_{l_1 n}), F(z_{l_2 n})|\widehat{\boldsymbol{\phi}}),
\end{aligned} \tag{2.18}$$

where q_c is the number of parameters in the copula model and N is the number of observations.

2.4 Efficient R-Vine Copula Based Fusion with Statistical Dependent Decisions

As observed in the optimal test statistic in Equation (2.5), the set of joint PMFs P_s and $Q_s, s \in S$ are required to be obtained at each time instant. To tackle the computational complexity resulting from multi-dimensional integration, we propose an efficient R-Vine copula based fusion approach of dependent decisions.

Let the local sensor probability of detection and local sensor probability of false alarm be represented by p_l and q_l for sensor $l, l = 1, 2, \ldots, L$. Therefore, p_l and q_l are given as

$$\begin{aligned} p_l &= \int_{\tau_l}^{+\infty} f(z_l|H_1)dz_l, \\ q_l &= \int_{\tau_l}^{+\infty} f(z_l|H_0)dz_l, \end{aligned} \tag{2.19}$$

where τ_l is the quantization threshold for sensor l. The local optimal sensor thresholds under the Neyman-Pearson criterion are obtained by solving the following problem:

$$\begin{aligned} \underset{\tau_l}{\text{maximize}} \quad & p_l, \\ \text{subject to} \quad & q_l \leq \beta_l, \end{aligned} \tag{2.20}$$

where β_l is the constraint on the local probability of false alarm for sensor l, p_l and q_l are given in Equation (2.19).

Consider the set of joint PMFs under hypothesis H_1, namely $P_s, s \in S$. Let $\tilde{A}_l = \{u_1 u_2 \ldots u_l \ldots u_L | u_l = 0\}$ and $\tilde{A}_l^c$ denote the complement of $\tilde{A}_l$ for $l = 1, 2, \ldots, L$. Note that the union of the sets $\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_L$ is S. For the three-sensor case, we have $\tilde{A}_1 = \{\{011\}, \{010\}, \{001\}, \{000\}\}$, $\tilde{A}_2 = \{\{101\}, \{100\}, \{001\}, \{000\}\}$ and $\tilde{A}_3 = \{\{110\}, \{100\}, \{010\}, \{000\}\}$. For any $s \in S$,

the PMF under hypothesis H_1 is given as

$$P_s = P(\bigcap_{l=1}^{L} B_l), \tag{2.21}$$

where $B_l = \tilde{A}_l$ if $s_l = 0$, otherwise, $B_l = \tilde{A}_l^c$. P_s can be obtained using copula functions. For example, P_{101} is given as

$$\begin{aligned} P_{101} &= P(\tilde{A}_1^c \cap \tilde{A}_2 \cap \tilde{A}_3^c) \\ &= P(\tilde{A}_2 - \tilde{A}_2 \cap \tilde{A}_3 - \tilde{A}_1 \cap \tilde{A}_2 + \tilde{A}_1 \cap \tilde{A}_2 \cap \tilde{A}_3) \\ &= 1 - p_2 - C_{23}(1-p_2, 1-p_3) - C_{12}(1-p_1, 1-p_2) \\ &+ C_{123}(1-p_1, 1-p_2, 1-p_3), \end{aligned} \tag{2.22}$$

where C_{12}, C_{23} and C_{123} are copula functions.

Consider the three-sensor case, the joint PMFs under H_1 is given as

$$\begin{aligned} &P(u_1 = 0, u_2 = 0, u_3 = 0) = C_{123} \\ &P(u_1 = 0, u_2 = 0, u_3 = 1) = C_{12} - C_{123} \\ &P(u_1 = 0, u_2 = 1, u_3 = 0) = C_{13} - C_{123} \\ &P(u_1 = 0, u_2 = 1, u_3 = 1) = 1 - p_1 - C_{12} - C_{13} + C_{123} \\ &P(u_1 = 1, u_2 = 0, u_3 = 0) = C_{23} - C_{123} \\ &P(u_1 = 1, u_2 = 0, u_3 = 1) = 1 - p_2 - C_{12} - C_{23} + C_{123} \\ &P(u_1 = 1, u_2 = 1, u_3 = 0) = 1 - p_3 - C_{23} - C_{13} + C_{123} \end{aligned} \tag{2.23}$$

where we omit the marginal CDFs of C, namely $1 - p_l, l = 1, 2, \ldots, L$. Similarly, PMFs under H_0 are obtained with p_l replaced by q_l, $l = 1, 2, \ldots, L$.

Define $\mathcal{C}$ as the set that specifies all the copula functions involved in the PMFs of sensor

Algorithm 2.1 Sequential method to obtain the joint PDF of sensor observations.

Inputs: Marginal PDFs of local sensor observations $f(z_i|H_1)$ for sensor i, $i = 1, 2, \ldots, m, m \in [1, 2, \ldots, L]$, data $(z_{1n}, \ldots, z_{mn}), n = 1, 2, \ldots, N$ and a predefined copula library $\mathcal{C}$.
Output: Joint PDF of sensor observations.

1. Get marginal CDFs of local sensor observations F_i, $i = 1, 2, \ldots, m$.
2. Calculate the weight $\mathbf{w}_{i,j}$ for all possible pairs of sensors $\{i, j\}$, $1 \leq i \leq j \leq m$.
3. Select the maximum spanning tree that maximizes the sum of absolute empirical weights, i.e.,
$$T_1 = \max \sum_{e=\{i(e),j(e)\} \text{ in spanning tree}} |w_{i(e),j(e)}|.$$
4. For each edge $e \in E_1$, select a copula $C^*_{i(e),j(e)}$ and estimate the corresponding parameter(s) $\boldsymbol{\phi}^*_{i(e),j(e)}$.
5. Obtain $F_{i(e)|j(e)}(\mathbf{z}_{i(e)}|\mathbf{z}_{j(e)})$ and $F_{j(e)|i(e)}(\mathbf{z}_{j(e)}|\mathbf{z}_{i(e)})$ using Equation (1.12).
6. **For** $s = 2, \ldots, m-1$ **do**
   (a) Calculate the weight $w_{i(e),j(e)|D(e)}$ for all conditional variable pairs $\{i(e), j(e)|D(e)\}$ that can be part of T_s.
   (b) Among these edges, select the maximum spanning tree, i.e.,
$$T_s = \max \sum_{e=\{i(e),j(e)|D(e)\} \text{ in spanning tree}} |w_{i(e),j(e)|D(e)}|.$$
   (c) For each edge $e \in E_s$, select a best conditional copula $C^*_{i(e),j(e)|D(e)}$ and estimate the corresponding parameters $\boldsymbol{\phi}^*_{i(e),j(e)|D(e)}$.
   (d) Obtain $F_{i(e)|j(e)\cup D(e)}(\mathbf{z}_{i(e)}|\mathbf{z}_{j(e)}, \mathbf{z}_{D(e)})$ and $F_{j(e)|i(e)\cup D(e)}(\mathbf{z}_{j(e)}|\mathbf{z}_{i(e)}, \mathbf{z}_{D(e)})$ using Equation (1.12).
7. **end For**
8. Obtain the R-Vine copula density c.
9. Obtain the joint PDF of sensor observations using Equation (2.15).

decisions. We further define the index set of $\mathcal{C}$ as G which is the union of all the nonempty subsets with at least two elements of set $\{1, 2, \ldots, L\}$ in sorted order and the cardinality of set G is $|G| = N_G = \sum_{k=2}^{L} \binom{L}{k}$. For the three-sensor case, we have the copula function set $\mathcal{C} = \{C_{12}, C_{13}, C_{23}, C_{123}\}$ and its index set $G = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$.

As we can see, knowing $\mathcal{C}$, we can obtain all combinations of the joint PMFs. Any arbitrary copula density function of $C \in \mathcal{C}$ can be obtained through Algorithm 2.1. By integrating the copula density function, we can obtain the copula function $C \in \mathcal{C}$. The computation is significantly reduced using the copula function set $\mathcal{C}$ to obtain the joint PMFs since we only need to perform multi-dimensional integration once for each copula function $C \in \mathcal{C}$. To further reduce computational complexity, we start with L-dimensional R-Vine copula model selection by applying Algorithm 2.1 and then use the obtained optimal tree structure with its R-Vine matrix $\mathbf{M}^*$ (see Definition 1.3), R-Vine copula family matrix $\mathbf{F}^*$ and the corresponding parameter matrix P^* to directly get the copula density functions that need to be estimated in $\mathcal{C}$. For the rest of the copula functions to be estimated, we again start with selecting an appropriate R-Vine copula model with largest dimension and use its optimal tree structure to obtain lower dimensional copula functions that have not been estimated. We proceed with this procedure till we obtain all the copula functions in the set $\mathcal{C}$. For the R-Vine copula example in Fig. 1.1, from its R-Vine matrix $\mathbf{M}^*$ (see Section 1.1.5) with its optimal R-Vine copula family matrix and the corresponding parameter matrix, we can directly obtain the density of $c_{35}, c_{24}, c_{12}, c_{23}, c_{123}, c_{1234}, c_{12345}$.

The proposed efficient optimal fusion rule is summarized in Algorithm 2.2.

2.5 Simulation Results

In this section, we demonstrate the efficacy of our proposed R-Vine copula based fusion methodology for the problem of distributed detection through numerical examples. We assume that there are two hypotheses, where H_1 denotes the presence of a signal s and H_0 indicates

Algorithm 2.2 Efficient optimal fusion rule.

Inputs: Marginal PDFs of local sensor observations $f(\mathbf{z}_l|H_1)$, $l = 1, 2, \ldots, L$.
Output: Log optimal detection statistics.

1. Obtain optimal local quantizer threshold $\tau_l, l = 1, 2, \ldots, L$ for all sensors by solving problem in Equation (2.20).
2. Calculate local sensor probability of detection p_l and probability of false alarm q_l for all sensors, $l = 1, 2, \ldots, L$.
3. Obtain optimal R-Vine structure of L sensors using algorithm 2.1 and its R-Vine matrix $\mathbf{M}^*$ and the corresponding R-Vine copula family matrix $\mathbf{F}^*$ and parameter matrix $\mathbf{P}^*$.
4. For $i = 1, 2, \ldots, L-1$,
   (a) Let $G_1 = \emptyset$.
   (b) Obtain $C_{\mathbf{M}^*_{i,i},\mathbf{M}^*_{L,i}}$ and $C_{\mathbf{M}^*_{i,i},\mathbf{M}^*_{i+1,i},\ldots,\mathcal{M}^*_{L,i}}$ directly from the obtained R-Vine copula family matrix $\mathbf{F}^*$ and parameter matrix $\mathbf{P}^*$.
   (c) $G_1 = G_1 \cup \{\{\mathbf{M}^*_{i,i}, \mathbf{M}^*_{L,i}\}, \{\mathbf{M}^*_{i,i}, \mathbf{M}^*_{i+1,i}, \ldots, \mathbf{M}^*_{L,i}\}\}$.
5. For $g = 1, 2, \ldots, N_G - 1$,
   (a) if $G(g) \neq a, \forall a \in G_1$.
   (b) Apply algorithm 2.1 and obtain $C_{G(g)}$.
6. Calculate the PMFs of sensor decisions under hypotheses H_1 and H_0, respectively, using Equation (2.21).
7. Solve the detection testing problem in Equation (2.6).

the absence of s. In the distributed sensor network we consider in this chapter, we assume that three sensors sense and acquire raw measurements of the signal s via a linear sensing model, and then quantize the detected signal into a single-bit local decision. After compression, the decisions are transmitted to the FC. The signals received at the sensors can be modeled as:

$$\begin{aligned} H_1 &: z_{in} = h_{in}s_{in} + w_{in}, \qquad i = 1,2,3; n = 1,\ldots,N \\ H_0 &: z_{in} = w_{in}, \qquad i = 1,2,3; n = 1,\ldots,N \end{aligned} \tag{2.24}$$

where z_{in}, h_{in} and w_{in} denote the received signal, the fading channel gain and the measurement noise at sensor i and time instant n. Moreover, s_{in} is the target signal received by the ith sensor at nth time instant. The intensity of the signal s is assumed to be a constant. We assume that the channel gain h_{in} is chosen randomly and independently from Rayleigh(ξ) distribution with parameter ξ over time. However, h_{in} can be spatially dependent. The measurement noise w_{in} is drawn from zero-mean Gaussian distribution with standard deviation σ_w ($\sigma_{w1} = 1$, $\sigma_{w2} = 0.9$ and $\sigma_{w3} = 0.8$) and is assumed to be temporally independent conditioned on either hypothesis but can be spatially dependent. Furthermore, we assume that the measurement noise, the fading gains, and the target signal are mutually independent. Also, we assume that we do not have any prior knowledge of the marginals and dependence structure. Unless specified otherwise, the number of sensor observations is assumed to be $N = 100$, the local probability of false alarm is constrained by $q_l \leq 0.1, l = 1,2,3$ and AIC is used for optimal bivariate copula selection.

To demonstrate the superiority of R-Vine copula, we apply different multivariate copulas and seven different R-Vine copulas given by

1. Mixed R-Vine: R-Vine with pair-copula terms chosen individually from 15 bivariate copula types (Gauss, Student-t, Gumbel, Clayton, Frank and Joe etc.).

2. all Gaussian R-Vine: R-Vine with each pair-copula term chosen as bivariate Gaussian copula.

3. all Student t R-Vine: R-Vine with each pair-copula term chosen as bivariate Student t copula.

4. all Gumbel R-Vine: R-Vine with each pair-copula term chosen as bivariate Gumbel copula.

5. all Clayton R-Vine: R-Vine with each pair-copula term chosen as bivariate Clayton copula.

6. all Frank R-Vine: R-Vine with each pair-copula term chosen as bivariate Frank copula.

7. all Joe R-Vine: R-Vine with each pair-copula term chosen as bivariate Joe copula.

	MLE	AIC	BIC	p-value
R-Vine mixed	6300.72	-12595.44	-12575.88	0.92
R-Vine all Gaussian	4572.36	-9138.72	-9119.16	0.48
R-Vine all Student t	4868.76	-9725.52	-9686.42	0.38
R-Vine all Gumbel	5799.94	-11593.87	-11574.32	0.57
R-Vine all Clayton	6161.90	-12317.8	-12298.25	0.82
R-Vine all Frank	4553.14	-9100.29	-9080.74	0.57
R-Vine all Joe	6130.61	-12255.22	-12235.67	0.74
Multi-Clayton copula				0.0005
Multi-Gaussian copula				0.0005
Multi-Frank copula				0.0005

Table 2.1: The performance of R-Vine classes and standard multivariate copulas.

Performing a parametric bootstrap with repetition rate $B = 1000$ and sample size $N = 5000$, the goodness-of-fit test results are shown in Table 2.1, where the global MLE, AIC and BIC values are obtained by adding all the bivariate copula information scores calculated using Equation (2.18). As we can see, the p-value confirms that the R-Vine mixed model (the optimal fusion methodology) can not be rejected at a 5% significance level, i.e., that the R-Vine mixed model fits the data quite well. The R-Vine models with a single type of bivariate copulas have a smaller significance than the R-Vine mixed model. The standard multivariate copulas, e.g., multivariate Clayton, Gaussian and Frank copulas, are rejected at a 5% significance level. This

indicates that the standard multivariate copulas are quite limited in their ability to characterize complex dependence.

To exhibit the performance improvement by applying R-Vine copula based fusion of dependent sensor decisions, we also evaluate the detection performance obtained by using the Chair-Varshney fusion rule that assumes independence of sensor decisions. Here, the R-Vine copula based fusion rule is obtained by choosing from 40 bivariate copula types. We use receiver operating characteristics (ROCs) to characterize the detection performance. For clarity, we summarize the empirically studied cases as follows.

- Case 1: We assume that the fading channel gains are spatially dependent. The measurement noises and the target signals received at the local sensors are assumed to be spatially and temporally independent.

- Case 2: We assume that the target signals received at the local sensors are spatially dependent but are assumed to be temporally independent conditioned on either hypothesis. The measurement noises are assumed to be spatially and temporally independent. To characterize the performance of this case, we further assume that the channels are ideal.

- Case 3: We assume that the measurement noises are spatially dependent. The target signals received at the local sensors are assumed to be spatially and temporally independent and the channels are ideal .

In Fig. 2.1, we present the ROCs comparing the two fusion rules: the Chair-Varshney fusion rule and the proposed R-Vine copula based fusion rule for case 1 with different fading parameters, ξ. The intensity of the signal at the local sensors is assumed to be $s_i = 4, i = 1, 2, 3$. As we can see, the detection performance of the R-Vine copula based fusion rule is significantly better than that of the Chair-Varshney fusion rule. Moreover, with stronger fading ($\xi = 0.9$), we can see that the detection performance is degraded compared to the fading with parameter $\xi = 1$.

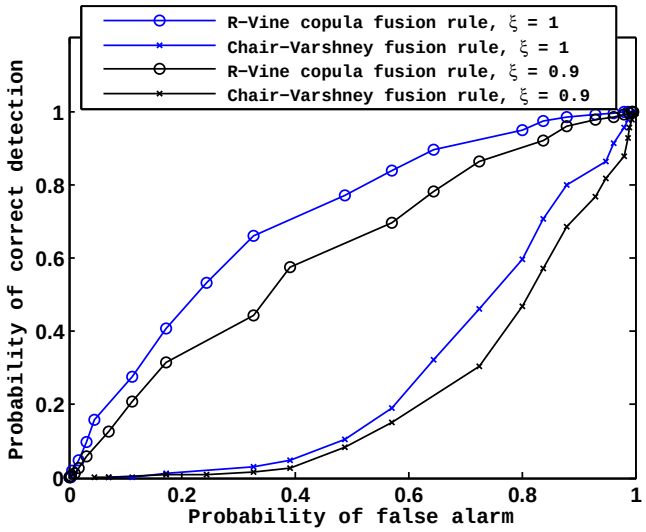

Fig. 2.1: ROCs comparing the Chair-Varshney fusion rule and the R-Vine copula based fusion rule with dependent fading channels.

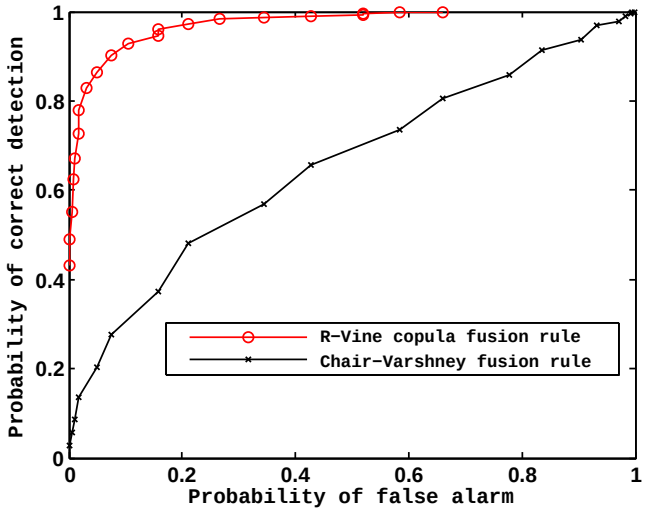

Fig. 2.2: ROCs comparing the Chair-Varshney fusion rule and the R-Vine copula based fusion rule with dependent signals.

In Fig. 2.2 and Fig. 2.3, we give the ROCs comparing the Chair-Varshney fusion rule and the proposed R-Vine copula based fusion rule for Case 2 under different dependence structures. The intensity of the signal received at the local sensors is assumed to be $s_i = 2.4, i = 1, 2, 3$. Fig. 2.2 shows the detection performance under a strong dependence structure and Fig. 2.3 gives the detection performance under a weaker dependence structure. As we can see, for both

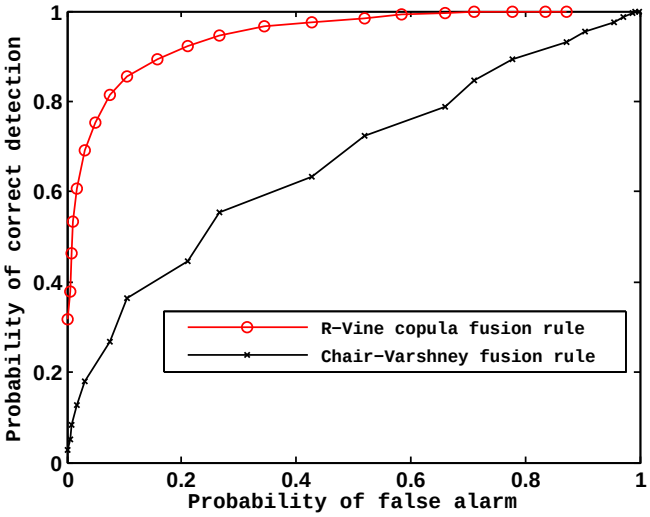

Fig. 2.3: ROCs comparing the Chair-Varshney fusion rule and the R-Vine copula based fusion rule with dependent signals for weaker dependence.

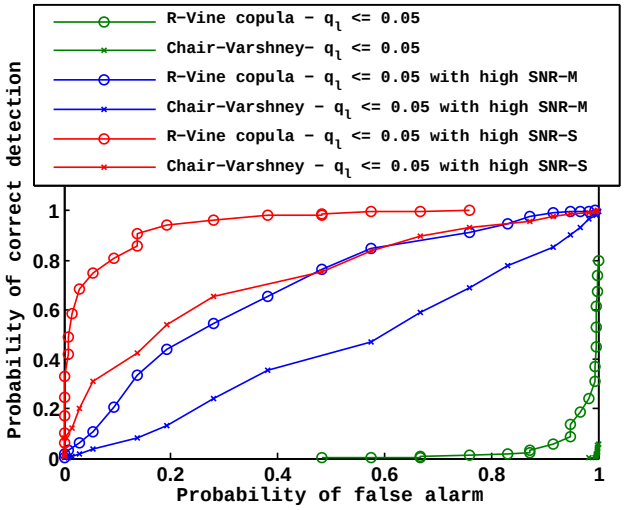

Fig. 2.4: ROCs comparing the Chair-Varshney fusion rule and the R-Vine copula based fusion rule with dependent signals for $q_l \leq 0.05$.

scenarios, the detection performance of the R-Vine copula based fusion rule is significantly better than that of the Chair-Varshney fusion rule. We further show the ROCs with the local probability of false alarm constrained by $q_l \leq 0.05, l = 1, 2, 3$ in Fig. 2.4. We can see that it is very difficult to detect the presence of the target signal for both the fusion rules as we have more tight false alarm constraints. By increasing the intensity of the signal to be $s_i =$

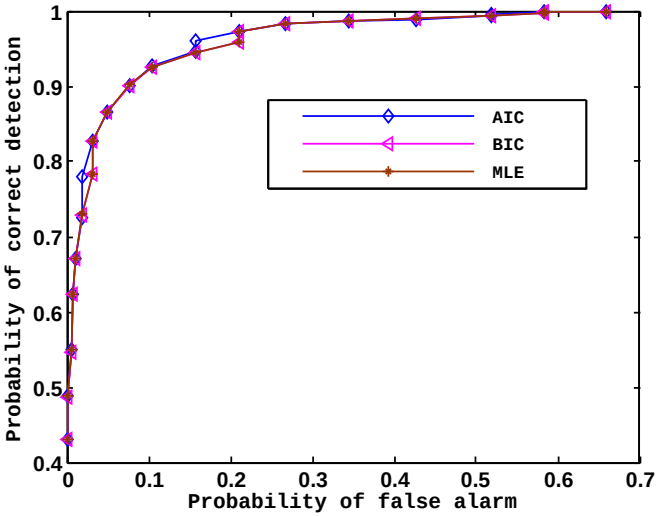

Fig. 2.5: ROCs for R-Vine copula based fusion rule with dependent signals for three model selection criteria.

$3, i = 1, 2, 3$, namely with high signal to noise ratio (SNR) in terms of stronger signal power (denoted by SNR-S) or decreasing the standard deviation of the measurement noise to be $\sigma_{wi} = 0.7, i = 1, 2, 3$, namely with high SNR in terms of weaker measurement noise power (denoted by SNR-M), we can see that the detection performance is much better compared to weaker signal intensity or stronger measurement noise cases.

In Fig. 2.5, we show the ROCs comparing the different model selection criteria discussed in Section 2.3.3, namely, AIC, BIC and MLE for the proposed R-Vine copula based fusion rule. As we observe, all the three criteria perform very well. The AIC criterion performs slightly better than the BIC and MLE criteria.

In Fig. 2.6, we present the ROCs comparing the Chair-Varshney fusion rule and the proposed R-Vine copula based fusion rule for Case 3. As expected, the detection performance of the R-Vine copula based fusion rule is much superior to that of the Chair-Varshney fusion rule.

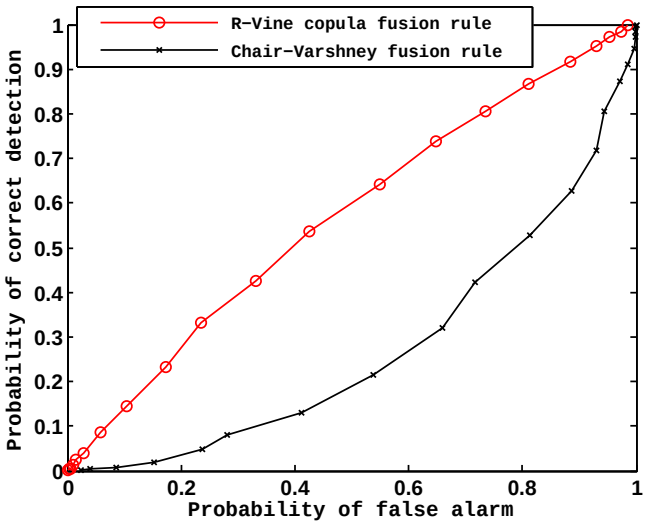

Fig. 2.6: ROCs comparing the Chair-Varshney fusion rule and the R-Vine copula based fusion rule with dependent measurement noise.

2.6 Summary

In this chapter, we studied the problem of distributed detection with dependent sensor decisions. We proposed a novel and powerful methodology to fuse dependent decisions obtained by binary quantization of statistically dependent sensor observations under the Neyman-Pearson framework. To derive the optimal fusion rule, we used the R-Vine copula model to characterize the complex dependence among multiple sensors. The proposed R-Vine copula based fusion methodology was employed to overcome the limitation of the existing standard multivariate copulas, and since this methodology is extremely flexible to model complex dependence structures. The optimal log likelihood test statistics at the FC involves multi-dimensional integration at each time, leading to very high computational complexity. We proposed an efficient R-Vine copula based optimal fusion algorithm. Numerical results have illustrated the efficiency of our approach.

CHAPTER 3
COPULA BASED DISTRIBUTED PARALLEL COMPUTING PLATFORM

3.1 Motivation

Fusion and inference from heterogeneous data streams have to deal with the challenge of achieving efficiency both in terms of accuracy and processing time. In terms of inference accuracy, the underlying dependence of observations needs to be taken into account. Also, in terms of inference processing time, the learning process from data streams often leads to long response time, especially when more accurate and complex dependence modeling approaches (such as copula theory) are used. There is a severe lack of approaches that can provide fast and accurate solutions to inference problems based on fusion of heterogenous streaming data due to the fact that it is quite challenging.

To overcome this challenge, we propose a novel parallel platform, C-Storm (Copula-based Storm), for heterogeneous stream data fusion based on Storm (see Section 3.2 for details of Storm) and the copula-based dependence modeling approach. The novel marriage of copula theory and Storm's parallel architecture addresses the lack of approaches for efficiently tak-

ing both inference accuracy and processing time into consideration for heterogeneous data stream fusion. C-Storm has the following desirable features: 1) C-Storm offers fast inference responses. 2) C-Storm provides high inference accuracies. 3) C-Storm is a general-purpose inference platform that can support data fusion applications. 4) C-Storm is easy to use and its users do not need to know deep knowledge of Storm or copula theory.

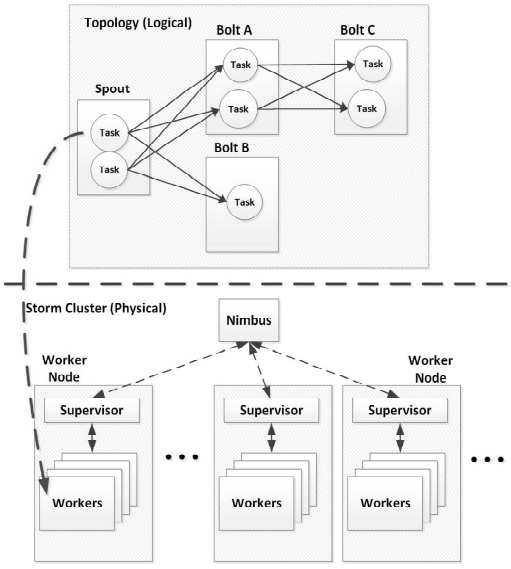

Fig. 3.1: The architecture of Storm.

3.2 Storm

Apache Storm [91] is a reliable and efficient computing platform for distributed/parallel stream data processing. The architecture of Storm is illustrated in Fig. 3.1. Storm uses two levels of abstractions, logical and physical, to express parallelism. In the physical layer (usually a cluster), there is a master node (known as *Nimbus*) and multiple worker nodes. The master node works as a central control unit to manage data processing that is actually done on worker nodes (i.e., physical or virtual machines). Each worker node runs a daemon called *superviser*

that listens to all the work assigned to it, on which *workers* process data tuples. In the logical layer, a directed graph known as topology is used to model an application, which includes two kinds of components: *spouts* and *bolts*. A spout is the source of a data stream, where data usually from external sources are read and emitted into the topology in the form of tuples. A bolt consumes tuples from any spout(s) or other bolt(s) and processes them based on user-defined functions. A bolt may emit new tuples and transmit them to the other bolts. The links in the topology graph indicate how tuples are routed. A spout or bolt can be executed as many tasks in parallel at runtime on one or multiple workers, which can be hosted by one or multiple worker nodes. A user can specify parallelism by configuring the number of workers for each component on a topology.

3.3 Design of C-Storm

In this section, we present the architecture and design details of C-Storm.

3.3.1 Architecture of C-Storm

As mentioned above, we develop C-Storm based on copula theory and Storm to embrace their power of dependance modeling and parallel data processing. The fusion rule for multi-sensor data requires complete knowledge of the form and structure of the joint distribution of sensor observations. The dependence structure of heterogeneous sources can be quite complex and nonlinear. Therefore, given arbitrary marginal distributions, their joint distribution cannot be simply written as the product of the marginals distributions. A major advantage of using copula-based dependence modeling approach (besides high inference accuracy for data with complex dependence) on Storm is that it allows separation of learning marginals (PDFs and CDFs) and learning the dependence structure (copula), i.e., they can be learned in parallel.

Suppose that a phenomenon or a target is continuously monitored by sensors $S_1, S_2, \ldots, S_L$.

Observations of sensor S_l are denoted by $\mathbf{z}_l = [z_{l1}, z_{l2}, \ldots, z_{lN}]^T$ $(l = 1, 2, \ldots, L)$. By fusing observations from all the sensors, we aim to achieve an inference (detection, classification, or estimation) goal. Using copula theory, a multivariate joint PDF is modeled as the product of marginal PDFs and the copula density (see Equation (1.2)).

Typically in many applications, we do not have any prior information related to the phenomenon of interest. Before designing the fusion rule, we need to determine the joint distribution of multivariate sensor observations, namely the marginal PDFs and the optimal copula density function. To obtain the optimal copula density function, we need to first have the knowledge of marginal CDFs (see Sklar's theorem in Section 1.1.1).

C-Storm enables the estimation of marginals (PDFs and CDFs) in a parallel way. Upon obtaining the estimated marginal CDFs, the best copula is selected in the following way: by estimating the copula parameters in parallel for each copula function in the library $\mathcal{C} = \{c_m : m = 1, \ldots, M\}$ and obtaining the likelihoods for each copula function, the best copula is selected to be the copula with maximum likelihood. Using the joint PDF we learn, fusion rules can be designed according to the problems we are interested in.

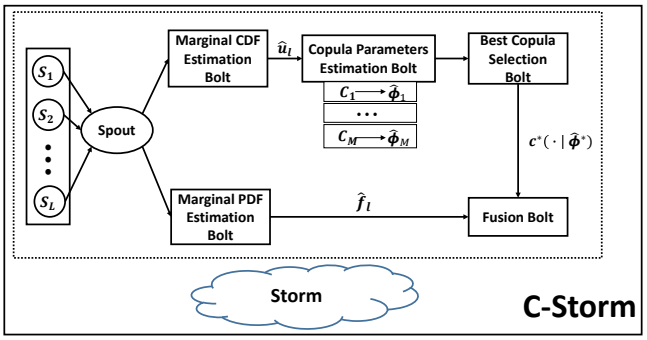

Fig. 3.2: The architecture of C-Storm.

The architecture for C-Storm is shown in Fig. 3.2, which can be considered as a combination of the original Storm and an additional software layer that runs as a Storm topology and implements the copula-based dependance modeling approach. C-Storm consists of a spout and 5 bolts as shown in the figure, whose functions are explained in the following:

- Spout: It keeps reading observations from the L sensors and emits tuples to marginal (PDF and CDF) estimation bolts.
- Marginal PDF Estimation bolt: Based on sensor observations, it provides a nonparametric and smoothed estimation of true density.
- Marginal CDF Estimation bolt: Based on sensor observations, it provides marginal CDFs needed to estimate the optimal copula (see Equation (1.1)).
- Copula Parameter Estimation bolt: Based on the marginal CDFs, it provides parameters of all the copula functions in the copula library.
- Best Copula Selection bolt: Based on all the copula functions and their parameters, it selects the best dependence structure to characterize the joint distributions of sensor observations.
- Fusion bolt: Based on the joint distributions of sensor observations, it uses a fusion rule (specified by the user), such as a log likelihood ratio test (for a detection problem), for inference according to the inference goal.

Our design has the following benefits:

- Fast Inference Responses: C-Storm offers fast inference responses since it leverages Storm's power of parallelism for fusing heterogeneous data streams by learning required marginals and dependence structure and performing fusion in parallel.
- High Inference Accuracies: C-Storm provides high inference accuracies since it uses the copula-based dependence modeling approach that can significantly improve inference accuracy compared to the commonly used independence modeling method.
- General-Purpose Inference Platform: C-Storm is a general-purpose inference platform that can support various data fusion applications since users can specify different fusion rules based on their application needs.

- Easy to Use: C-Storm is easy to use since a user only needs to provide data and specify the fusion rule without knowing the details of Storm or copula theory.

We explain the copula-based dependence modeling approach in some detail and summarize workflow of the proposed approach in the next subsection.

3.3.2 Copula-based Dependence Modeling

First, we explain how to estimate marginal PDFs and CDFs. The joint PDF of $\mathbf{z}_l$, $l = 1, 2, \ldots, L$ is given by

$$f(\mathbf{z}_1, \ldots, \mathbf{z}_L) = \prod_{n=1}^{N} \Big(\prod_{l=1}^{L} f(z_{ln}) \Big) c(u_{1n}, \ldots, u_{Ln} | \boldsymbol{\phi}), \tag{3.1}$$

where $f(z_{ln})$ is the marginal PDF, u_{ln} is the marginal CDF for sensor l at time instant n ($l = 1, 2, \ldots, L$, $n = 1, 2, \ldots, N$), and $\boldsymbol{\phi}$ is the parameter of copula c.

Without any prior knowledge of the phenomenon of interest, the marginal PDFs $f_l(\cdot)$ for sensor l ($l = 1, 2, \ldots, L$) can be estimated non-parametrically, and the marginal CDFs $\mathbf{u}_l = [u_{l1}, u_{l2}, \ldots, u_{lN}]$ ($l = 1, 2, \ldots, L$) can be determined by the EPIT in [45]. The estimate of uniform random variables u_{ln} is obtained by using EPIT:

$$\hat{F}(\cdot) = \frac{1}{N} \sum_{n=1}^{N+1} \mathbf{1}_{z_{ln} < \{\cdot\}}, \tag{3.2}$$

$$\hat{u}_{ln} = \hat{F}(z_{ln}), \tag{3.3}$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function.

Kernel density estimators [108] provide a smoothed estimate of true density by choosing the optimal bandwidth so that an accurate estimate is achieved. Leave-one-out cross-validation method is applied to choose the kernel bandwidth. For a kernel K, the optimal bandwidth h^* is obtained by minimizing the cross-validation estimator of risk $\hat{J}$. The risk estimator can be

easily acquired using the approximation in [108, p. 136]:

$$\hat{J}(h) = \frac{1}{hN^2} \sum_{p} \sum_{q} K^* \left(\frac{X_p - X_q}{h} \right) + \frac{2}{Nh} K(0) + \mathcal{O}\left(\frac{1}{N^2} \right), \tag{3.4}$$

where $K^*(x) = K^{(2)}(x) - 2K(x)$ and $K^{(2)}(z) = \int K(z-y)K(y)dy$.

Next, we discuss the estimation of copula parameters and the selection of the best copula. Before selecting the optimal copula, the copula parameter set $\boldsymbol{\phi}$ is obtained using MLE, which is given by

$$\widehat{\boldsymbol{\phi}} = \arg\max_{\boldsymbol{\phi}} \sum_{n=1}^{N} \log c(\hat{u}_{1n}, \ldots, \hat{u}_{Ln} | \boldsymbol{\phi}), \tag{3.5}$$

where $\hat{u}_{ln}$ is the empirical estimate of u_{ln}.

The best copula c^* (maximum likelihood) is selected from a predefined library of copulas, $\mathcal{C} = \{c_m : m = 1, \ldots, M\}$. It is given as

$$c^* = \arg\max_{c_m \in \mathcal{C}} \sum_{n=1}^{N} \log c_m(\hat{u}_{1n}, \ldots, \hat{u}_{Ln} | \widehat{\boldsymbol{\phi}}_m). \tag{3.6}$$

In summary, C-Storm works as follows: first, the spout keeps reading observations from multiple sources and emits the corresponding tuples to the marginals estimation bolts, which estimate marginal CDFs and PDFs using Equation (3.3) and the kernel density estimators respectively. Second, the copula parameter estimation bolt estimates parameters of all the copula functions in the library $\mathcal{C}$ based on estimated marginal CDFs using Equation (3.5). Third, the best copula selection bolt outputs the optimal copula and its parameters based on all the copula functions and their parameters using Equation (3.6). In the last step, the fusion bolt uses the estimated marginal PDFs, the best copula and its parameters to achieve the given inference goal.

3.4 Simulation Results

For validation and performance evaluation, we implemented C-Storm based on Apache Storm 1.0.2 [91]. We conducted extensive experiments on a Storm cluster of 6 Ubuntu 14.04 VMs, each of which is equipped with a 2-core virtual CPU running at 2.30GHz and 2GB of RAM.

3.4.1 Fusion Application and Experimental Setup

In our experiments, we considered a detection problem, where a random phenomenon is monitored continuously by L sensors. A binary hypothesis testing problem is studied, where H_0 denotes the absence of the phenomenon (null hypothesis) and H_1 denotes the presence of the random phenomenon (alternative hypothesis). The lth sensor ($l = 1, 2, \ldots, L$) makes a set of N observations, $\mathbf{z}_l = [z_{l1}, z_{l2}, \ldots, z_{lN}]^T$. We assume that sensor observations are continuous random variables and i.i.d. over time. The collective raw observations, $\mathbf{z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_L]$, are transmitted to C-Storm. By estimating the joint distributions of $\mathbf{z}$, C-Storm determines whether a phenomenon is present or not. Since sensor observations are independent over time, the likelihood ratio test statistic is given as

$$\Lambda(\mathbf{z}) = \frac{\prod_{n=1}^{N} f_1(z_{1n}, z_{2n}, \ldots, z_{Ln}|H_1)}{\prod_{n=1}^{N} f_0(z_{1n}, z_{2n}, \ldots, z_{Ln}|H_0)}, \tag{3.7}$$

where f_1 and f_0 denote the joint PDFs under alternative and null hypotheses, respectively.

Using the copula-based dependence modeling approach and taking log on both sides of Equation (3.7), the log test statistic can be expressed in terms of the optimal copula densities, c_1^* and c_0^*, respectively under H_1 and H_0, as

$$\begin{aligned} \log \Lambda(\mathbf{z}) = &\sum_{n=1}^{N} \sum_{l=1}^{L} \log \frac{\widehat{f}_1(z_{ln})}{\widehat{f}_0(z_{ln})} \\ &+ \sum_{n=1}^{N} \log \frac{c_1^*(\hat{u}_{1n}^1, \ldots, \hat{u}_{Ln}^1|\widehat{\boldsymbol{\phi}}_1^*)}{c_0^*(\hat{u}_{1n}^0, \ldots, \hat{u}_{Ln}^0|\widehat{\boldsymbol{\phi}}_0^*)}, \end{aligned} \tag{3.8}$$

where $\widehat{f}_k(z_{ln})$ is the estimated marginal PDF, $\hat{u}_{ln}^k = F(z_{ln}|H_k)$ denotes the estimated CDF, for sensor l at time instant n and $\widehat{\boldsymbol{\phi}}_k^*$ is the parameter of the optimal copula c_k^*, under hypothesis H_k ($k = 0, 1$). The optimal fusion rule is given by

$$\log \Lambda(\mathbf{z}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{3.9}$$

where η is the threshold for the test.

With unknown marginals and dependence structure, the test statistic is given by Equation (3.8), where the marginal CDFs $\hat{u}_{ln}^k$, marginal PDFs $\widehat{f}_k(z_{ln})$ and optimal copulas c_k^* ($k = 0, 1$, $l = 1, 2, \ldots, L$, $n = 1, 2, \ldots, N$) need to be learned from the measurements $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_L$ within a short time. Since sensor observations arrive continuously at a high rate, the learning process should be fast enough so that incoming data can be processed in a timely manner.

For training, the spout and the bolts work as described in Section 3.3, i.e., C-Storm computes the marginal PDFs, the optimal copulas and its parameters under hypotheses H_1 and H_0 for hypothesis testing. After training, hypothesis testing starts. For hypothesis testing, we implemented a fusion bolt, which performs the fusion rule given by Equation (3.9) based on the training results.

We tested the detection topology with 6 worker nodes, each of which hosts only one worker. Our experimental results are presented for a 2-sensor case ($L = 2$). Note that the copula-based dependence modeling approach described above can be easily extended to the general case where $L > 2$ since one can construct a multivariate copula using bivariate components [92]. The input data were generated with normal and beta distributed marginals and Student t copula dependence [75]. Note that we did not include Student t copula in the copula library of the proposed approach for fair comparison.

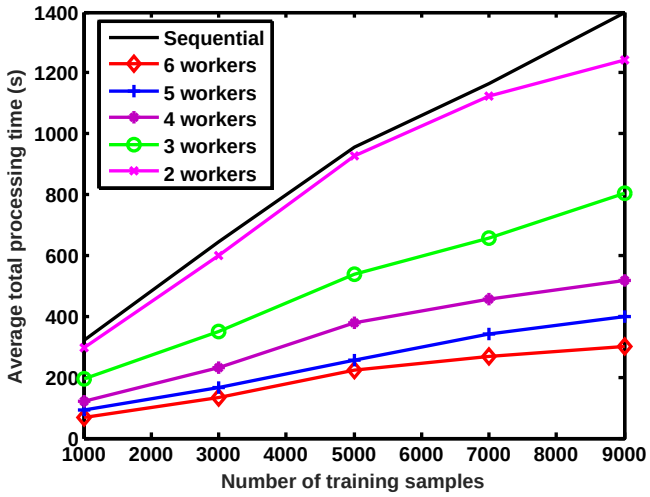

Fig. 3.3: C-Storm versus sequential baseline in terms of average total processing time.

3.4.2 Experimental Results and Analysis

To show the superiority of C-Storm, we compared it with a sequential baseline (labeled as "Sequential") using the same dependence modeling approach. For fair comparison, we implement and run the baseline on Storm using one worker. To reduce the effect of randomness, every number presented in the following figures is an average over 5 Monte Carlo runs. Fig. 3.3 shows the average total processing times achieved by C-Storm with $W \in \{2, 3, 4, 5, 6\}$ workers and the sequential baseline respectively for different training data sizes. From the results, we can see that C-Storm significantly accelerates data fusion and reduces processing time. For example, when the number of training data samples is 5000 and 4 workers are used, C-Storm achieves an average total processing time of 379s, while the average processing time given by the sequential baseline is 955.3s. C-Storm offers an average of 2.6x speedup for the case with 4 workers. When the number of workers becomes 6 and the number of training samples is 5000, C-Storm reduces the average total processing time from 955.3s (sequential) to 207.1s, which represents a speedup of 4.6x. On an average, C-Storm offers a 4.7x speedup with 6 workers. As expected, more workers, i.e., higher degree of parallelism, leads to more speedup and less processing time. The speedup when using 2 workers is not as impressive as that when

using more $(3-6)$ workers. This is because before workers start to work, Storm needs some time to distribute tasks to each worker and such overhead counteracts the expected speedup. In addition, it can also be seen that the average total processing time increases with the number of training data samples. In the fusion bolt described above, the marginal probability densities under both hypotheses are obtained first on testing data based on the estimated marginal PDFs, and the copula densities under both hypotheses are obtained on testing data. Then, the test statistic starts to work. Fig. 3.4 shows the average total processing time of C-Storm with different number of workers used for obtaining the marginal probability densities on the testing data in the fusion bolt. As we can see, within 300s, C-Storm can process 5000 training data samples using 4 workers and 7000 training data samples using 6 workers. On an average, by using 6 workers, C-Storm can process 40% more data samples than using 4 workers.

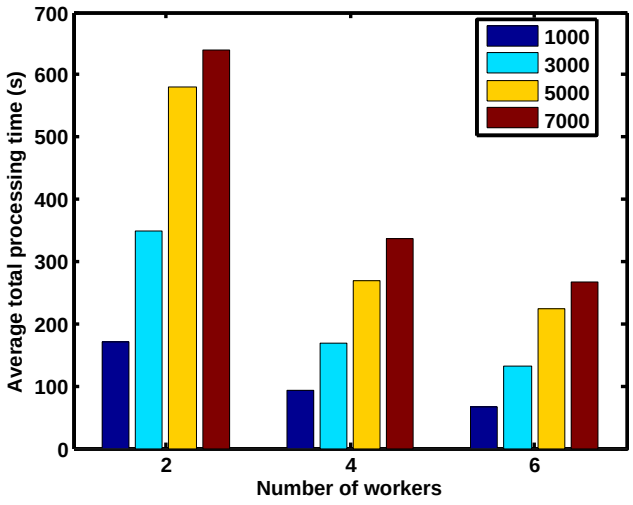

Fig. 3.4: Average total processing times of C-Storm with different number of workers used for obtaining the marginal probability densities in the fusion bolt.

Fig. 3.5 shows the average total processing time of C-Storm with different number of workers used for performing the test given by Equation (3.8) in the fusion bolt. Here, we can make similar observations that more workers can process more data samples. The results from these two experiments are consistent with those in the first experiment, which again validate our claim that higher degree parallelism leads to less processing time for stream data fusion.

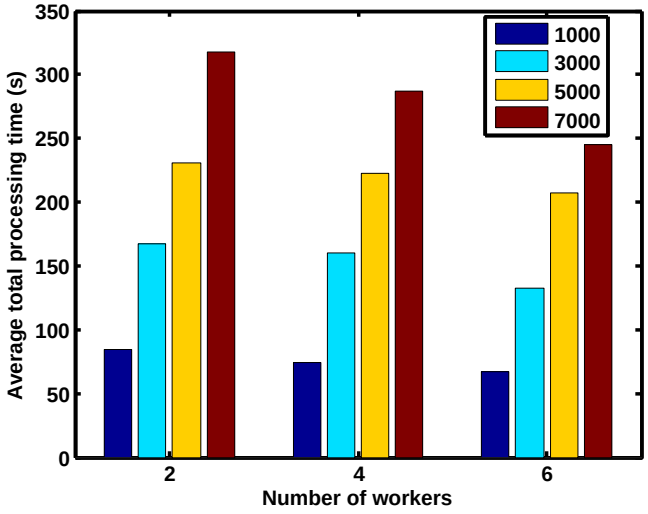

Fig. 3.5: Average total processing times of C-Storm with different number of workers used for performing the test in the fusion bolt.

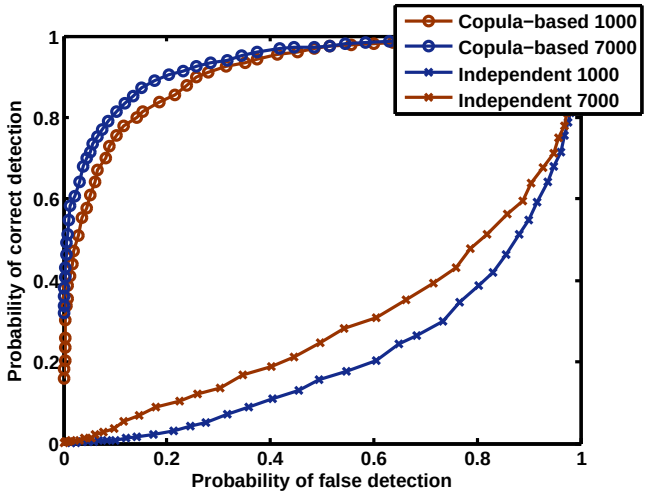

Fig. 3.6: ROCs for the detection problem.

As mentioned above, we are concerned about both processing time and detection accuracy for data fusion applications. Fig. 3.6 presents the detection accuracy using the Receiver Operating Characteristics (ROC) for copula-based dependence modeling and independence modeling with different training data sizes. We can observe that copula-based dependence modeling significantly improves detection accuracy compared to the independence modeling approach that assumes that sensor observations are independent.

3.5 Summary

In this chapter, we presented the design and evaluation of C-Storm, which is a novel parallel platform that is built based on Apache Storm and uses the copula-based dependence modeling approach for the fusion of heterogeneous data streams. C-Storm offers fast inference responses and high inference accuracies. Moreover, it is a general and easy-to-use platform that can support various data fusion applications. Its users do not need to know the details of Storm or copula theory. We demonstrated the superiority of C-Storm via a detection application. Experimental results have shown that C-Storm achieves 4.7x speedup over a sequential baseline on average, and higher degree of parallelism leads to better performance.

CHAPTER 4
DISTRIBUTED CLASSIFICATION WITH DEPENDENT FEATURES

4.1 Motivation

Human activity recognition (HAR) is quite important as it can be used for health care, personal fitness, and border surveillance, etc. [49, 110, 111]. The task of HAR is to detect and recognize human actions from the data provided by multiple sensors. HAR is naturally a classification task. Combining multiple sensing modalities can boost the classification performance. However, since each sensor carries a unique physical trait, sensor heterogeneity or incommensurability is the first critical challenge for multi-modal fusion. Also, multiple sensor modalities tend to be dependent due to non-linear cross-modal interactions.

Most of the current multi-classifier fusion solutions for HAR rely on *shallow* classifiers, such as Support Vector Machines (SVMs), Random Forests and Decision Trees, which employ handcrafted statistical features extracted from each modality. The typical strategy for the fusion of these features is to combine the outputs obtained from multiple classifiers, where each classifier only takes the features of one modality [9, 37, 64, 79, 89, 116]. Note that designing and

selecting robust features heavily relies on human experience and is time consuming. Also, only shallow features, such as mean, variance and amplitude, can be learned according to human expertise, which can be insufficient for more complex activities [114].

Very recently, multimodal deep learning methodologies for HAR have attracted some attention [74, 84]. Compared to the shallow classifiers, deep classifiers can learn many more high-level features directly from raw data (or lightly processed data) and avoid the need for the design of handcrafted features. Three fusion strategies can be applied to deep neural networks, based on the level where the fusion is performed: intermediate fusion with higher-level representations, referred to as high-level features, late fusion with decisions or late fusion with probability scores. In [74, 84], intermediate fusion strategies using Deep Neural Networks (DNNs) and Convolutional Neural Networks (CNNs) were studied, where a fully connected fusion layer was used to combine multiple DNNs or CNNs. As mentioned earlier, the data from multiple sensing modalities are non-linearly dependent. The fully connected fusion layer can learn this dependence in some manner. However, understanding and analyzing this non-linear dependence using the fully connected layer or another deep neural network is still an open question.

In this chapter, we leverage the DNNs and the R-Vine copula based dependence modeling for sensor-based recognition of human actions. More specifically, we use multiple DNNs to extract high-level features from multiple sensing modalities, where each DNN only processes the data from a single sensor. Thus, the data compatibility issues among multiple modalities can be avoided [62,76]. Different from the fusion strategy (using a fully connected fusion layer) in [74, 84], we propose a probabilistic fusion methodology, R-Vine copula based fusion rule, that combines the extracted high-level features and characterizes the cross-modal dependence. Moreover, our proposed model is designed to improve the classification performance compared to the neural network based fusion method and adds interpretability in the sense that it explicitly explains the dependence structure of the extracted features from different modalities.

4.2 Problem Formulation

Consider a supervised classification problem with G classes. Let $\Omega = \{w_1, w_2, \ldots, w_G\}$ be the set of class labels. L sensors make a set of observations regarding the object/event at time instant n, $\{\mathbf{x}_{1n}, \mathbf{x}_{2n}, \ldots, \mathbf{x}_{Ln}, y_n\}$, where $n = 1, 2, \ldots$ and $\mathbf{x}_{ln} \in \mathbb{R}^{d_l^1 \times d_l^2}, d_l^1, d_l^2 \in \mathbb{N}, \mathbb{N} = [1, 2, \ldots]$ is the observation of sensor l at time n. $y_n \in \Omega$ is the class label. We assume that the sensor observations are continuous random variables that are conditionally i.i.d. over time. L independent pre-trained DNN classifiers are used to extract high-level features from each sensing modality. A typical DNN is shown in Fig. 4.1. Compared to the traditional artificial neural networks, DNN is more capable of learning informative features from large amounts of data. We use $\mathbf{h}_n^l \in \mathbb{R}^{1 \times r_l}, r_l \in \mathbb{N}$ to represent the nth high-level feature vector extracted from sensor l. These high-level features are then combined using the R-Vine copula based fusion rule. We show the classification system studied here in Fig. 4.2.

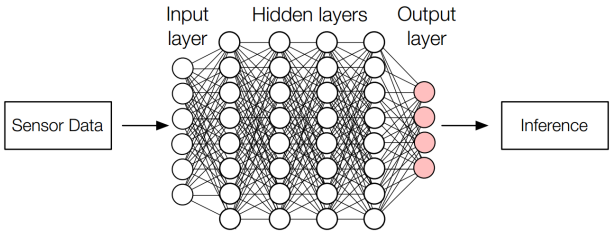

Fig. 4.1: A typical Deep Neural Network structure [84].

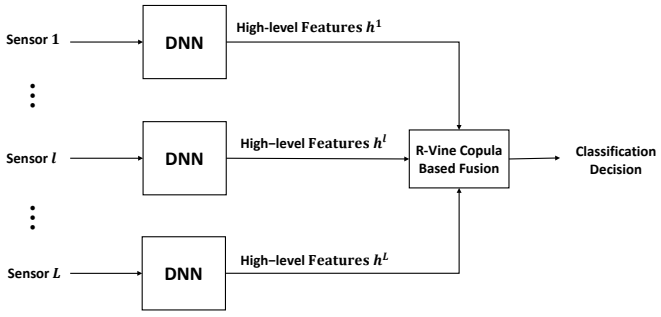

Fig. 4.2: R-Vine copula based multi-modal DNN.

Remark 4.1. *Note that for the sensor-based HAR, we use the feed-forward DNNs shown in Fig. 4.1 to extract high-level features instead of CNNs. There are two main reasons. The first one is that compared to DNNs, CNNs are computationally more intensive. The second one is that the high-level features extracted from CNNs are generally high dimensional. Also, among these high dimensional features, a large number of features are irrelevant and redundant. Fusing all the features based on R-Vine copula models is computationally inefficient.*

Our aim is to determine the class label by combining the extracted high-level features. Assume that we have a training set with a total of N feature vectors and the joint feature vector is

$$\mathbf{h}_n = [\mathbf{h}_n^1 \ \mathbf{h}_n^2 \ \ldots \ \mathbf{h}_n^L] \in \mathbb{R}^{1\times(r_1+r_2+\ldots+r_L)}, n = [N], \tag{4.1}$$

where $[N] = [1, 2, \ldots, N]$. In the following, for notational simplicity, we omit the superscripts of the feature vectors in Equation (4.1) and let $\mathbf{h}_n = [h_{1n}, h_{2n}, \ldots, h_{Kn}], K = r_1+r_2+\ldots+r_L$.

Using Bayes' theorem, the posterior probability of class w_i given the joint high-level feature vectors is given as:

$$P(w_i|\mathbf{h}) = \frac{f(\mathbf{h}|w_i)P(w_i)}{f(\mathbf{h})} \propto f(\mathbf{h}|w_i)P(w_i), \tag{4.2}$$

where $\mathbf{h} = [\mathbf{h}_1, \ldots, \mathbf{h}_N]$, $f(\mathbf{h}|w_i)$ is the joint likelihood function and $P(w_i)$ is the prior probability for class w_i. If the class prior probabilities are not known, it is commonly assumed that the classes are equally likely. The class label w_0 is determined by choosing the label with highest posterior probability, which is given by

$$w_0 = \arg\max_{w_i\in\Omega} P(w_i|\mathbf{h}). \tag{4.3}$$

Since $f(\mathbf{h})$ is a constant for all the classes, the main problem is how to model and maximize $f(\mathbf{h}|w_i)$ under unknown multivariate dependence. In the following section, we will use R-Vine copulas to model the joint likelihood function.

4.3 R-Vine Copula Based Fusion of Multiple Deep Neural Networks

In this section, we present the R-Vine copula based fusion rule. Our goal now is to find the joint PDFs of feature vectors $\mathbf{h}$ under each class. According to Sklar's theorem (Section 1.1.1), the joint PDF can be separated into its marginals and the dependence structure that is fully characterized by the copula (see Equation (1.2)). Therefore, we have

$$f(\mathbf{h}|w_i) = \prod_{n=1}^{N} \Big(\prod_{k=1}^{K} f_k(h_{kn}|w_i) \Big) c_i \Big(\mathbf{F}^i(\mathbf{h}_n) | \boldsymbol{\phi}_i \Big), \tag{4.4}$$

where $f_k(h_{kn}|w_i), k = 1, \ldots, K$ are the marginal PDFs and $\mathbf{F}^i(\mathbf{h}_n) = [F_1^i(h_{1n}), \ldots, F_K^i(h_{Kn})]$ denotes all the marginal CDFs at time instant n under class $w_i, w_i \in \Omega$. Moreover, c_i is the copula density function for class w_i and $\boldsymbol{\phi}_i$ is the corresponding parameter set.

Since we have no knowledge of the joint distributions of the extracted high-level features, the marginal PDFs, marginal CDFs, copula density functions and their corresponding parameters need to be estimated using the training dataset. The estimation of the marginal distributions and optimal copula density functions for all the classes is similar. Therefore, the class index i is omitted in the rest of the chapter.

The marginal PDFs can be estimated non-parametrically using kernel density estimators [108] that provide a smoothed estimate of true density by choosing the optimal bandwidth so that an accurate estimate is achieved. Further, the marginal CDFs can be determined by the EPIT. The estimate of $F_k(\cdot)$ is given as

$$\hat{F}_k(\cdot) = \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{h_{kn} < \{\cdot\}}, \tag{4.5}$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function.

Next, we discuss how to construct and find the optimal multivariate copula c^* using R-Vine

copula models, which was introduced by Bedford and Cooke in [11, 12].

4.3.1 R-Vine copula Models

Using the R-Vine copula model and Sklar's theorem, the joint PDF of the feature vector $\mathbf{h} = [\mathbf{h}_1, \ldots, \mathbf{h}_N]$ is given by

$$f(\mathbf{h}|w_i) = \prod_{n=1}^{N}\prod_{k=1}^{K} f_k(h_{kn}|w_i) \prod_{m=1}^{K-1}\prod_{e\in E_m} \times \tag{4.6}$$
$$c_{\mathfrak{C}_{e,a},\mathfrak{C}_{e,b}|D_e}(F_{\mathfrak{C}_{e,a}|D_e}(h_{\mathfrak{C}_{e,a}n}|\mathbf{h}_{D_e}), F_{\mathfrak{C}_{e,b}|D_e}(h_{\mathfrak{C}_{e,b}n}|\mathbf{h}_{D_e n}); \boldsymbol{\phi}),$$

where $e = \{a, b\}$ and $\mathbf{h}_{D_e n} = \{h_{jn}|j \in D_e\}$, $f_k(\cdot|w_i)$ is the marginal PDF for kth feature, $k = 1, \ldots, K$.

4.3.2 Estimation of Optimal R-Vine copula

The estimation of optimal R-Vine copula model for the joint feature vector $\mathbf{h}$ requires the selection of the R-Vine tree structure $\mathcal{V}$, the choice of copula families for the bivariate copula set $\mathbf{B}$ and the estimation of their corresponding parameters $\boldsymbol{\phi}$. To select the optimal R-Vine tree structure, we adopt the sequential maximum spanning tree algorithm in [27]. This sequential method is based on Kendall's τ. The sequential method starts with the selection of the first tree T_1 and continues tree by tree up to the last tree T_{K-1}. The trees are selected in a way that the chosen bivariate copula models the strongest pair-wise dependencies present which are characterized by Kendall's τ. After the optimal R-Vine tree structure is selected, we need to define a bivariate copula family and estimate the optimal bivariate copulas that best characterizes the pair-wise dependencies.

Consider a library of copula, $\mathcal{C} = \{c_m : m = 1, 2, \ldots, M\}$. Before estimating the optimal

bivariate copula, the copula parameter set $\boldsymbol{\phi}$ is first obtained using MLE, which is given by

$$\hat{\boldsymbol{\phi}} = \arg\max_{\boldsymbol{\phi}} \sum_{n=1}^{N} \log c(\hat{F}_{k_1}(h_{k_1 n}), \hat{F}_{k_2}(h_{k_2 n})|\boldsymbol{\phi}), \tag{4.7}$$

where $(k_1, k_2), k_1, k_2 \in [1, 2, \ldots, K]$ is a connected pair in the selected R-Vine tree $\mathcal{V}$ and for simplification of notation, we omit the conditioning elements for conditional marginal CDFs. Note that the conditional marginal CDFs can be obtained recursively using Equation (1.12).

The best copula c^* is selected from the copula library $\mathcal{C}$ using the Akaike Information Criterion (AIC) [3] as the criterion, which is given as

$$\text{AIC}_m = -\sum_{n=1}^{N} \log c_m(\hat{F}_{k_1}(h_{k_1 n}), \hat{F}_{k_2}(h_{k_2 n})|\hat{\boldsymbol{\phi}}_m) + 2q(K), \tag{4.8}$$

where $q(K)$ is the number of parameters in the mth copula model. Also, the conditioning elements for conditional marginal CDFs are omitted.

The best copula c^* is

$$c^* = \arg\min_{c_m \in \mathcal{C}} \text{AIC}_m. \tag{4.9}$$

4.4 Simulation Results

In this section, we demonstrate the efficacy of our proposed R-Vine copula based methodology for the fusion of multiple DNNs. To show the superiority of our proposed fusion scheme, we also compare our result with the classification performance obtained by using the following schemes:

- Single modality without fusion: Feed the raw data into a DNN classifier.
- Data-level fusion: Concatenate all the raw data from different modalities into one input vector and feed it into a DNN classifier.

- Fully connected layer fusion: Concatenate the extracted features into one feature vector and use a fully connected fusion layer to achieve a final classification decision.

4.4.1 Datasets

We select two publicly available datasets that contain multi-modality sensor readings for the recognition of human activities.

STISEN Heterogeneity Activity Recognition Dataset, collected by Stisen et al. [90], contains the sensor readings from two modalities: smartphone and smart watch. Each modality is equipped with two motion sensors, accelerometer and gyroscope. There are 6 classes ('Sit', 'Stand', 'Walk', 'Stairsup', 'Stairsdown', 'Bike') to be classified. We focus on the fusion of phone and watch modalities. Each of the two motion sensors produces a three-dimensional data vector, making each data sample contain 6 attributes in total. We select the data captured by Samsung Galaxy S3 mini phone and Samsung Galaxy Gear watch, where the data samples were sampled at a rate of 100 Hz. 9000 samples from each modality are used to train and test DNN models for feature selection, and another 9000 samples are used to train and test the R-Vine copula based fusion methodology.

ANGUITA Human Activity Recognition Using Smartphone Dataset, collected by Anguita et al. [5], contains accelerometer and gyroscope three-dimensional sensor data. It was collected from 30 volunteers who performed six different activities ('Walking', 'Walking-upstairs', 'Walking-downstairs', 'Siting', 'Standing', 'Laying'). We focus on the fusion of accelerometer and gyroscope modalities. These sensor data were sampled at a rate of 50 Hz, and were separated into windows of 128 values. Each window has 50% overlap with the previous window. The 128-real value vector in each window stands for one sample for each activity.

4.4.2 Classification Accuracy

We use F_1 score as the classification performance metric, which is given by

$$F_1 = \frac{2}{|\Omega|} \sum_{w} \frac{\text{precision}_w \times \text{recall}_w}{\text{precision}_w + \text{recall}_w}, \tag{4.10}$$

where precision $= \frac{TP}{TP+FP}$ and recall $= \frac{TP}{TP+FN}$. Here, TP, FP and FN denote true positive, false positive and false negative, respectively. F_1 score is robust to unbalanced distributions of data samples across classes.

Model	F_1 score
Watch-DNN	71.4%
Phone-DNN	70.2%
Fully-connected layer fusion	78.0%
Data-level fusion	79.3%
R-Vine copula fusion	88.6%

Table 4.1: **STISEN**: F_1 scores for Watch-DNN, Phone-DNN, Fully-connected layer fusion, Data-level fusion, R-Vine copula fusion.

Model	F_1 score
Accelerometer-DNN	87.8%
Gyroscope-DNN	72.9%
Fully-connected layer fusion	91.9%
Data-level fusion	88.3%
R-Vine copula fusion	92.8%

Table 4.2: **ANGUITA**: F_1 scores for Accelerometer-DNN, Gyroscope-DNN, Fully-connected layer fusion, Data-level fusion, R-Vine copula fusion.

Table 4.1 and Table 4.2 show the F_1 scores comparing the five classification schemes: two single modality based DNN classifiers and three multi-modal fusion based DNN classifiers for the **STISEN** and **ANGUITA** datasets, respectively. As we can see, fusion based schemes perform better than single modality based schemes. Also, our proposed R-Vine copula based fusion methodology performs better than using the data-level fusion scheme and fully connected fusion layer scheme. Our proposed methodology achieves an overall 88.6% and 92.8%

F_1 scores for the **STISEN** and **ANGUITA** datasets, respectively. Moreover, we can see that for phone and watch modalities, the R-Vine copula based fusion scheme achieves higher performance improvement compared to accelerometer and gyroscope modalities. This is because of the fact that the accelerometer and gyroscope are less dependent while the phone and watch are highly dependent.

It should be noted that the training of R-Vine copula models requires less number of training samples compared to the training of a fully connected fusion layer or another DNN used for fusion. In Table 4.1 and Table 4.2, the performance of our R-Vine copula scheme is obtained by using a total of $N = 1200$ feature samples. However, the fully connected fusion layer based scheme requires a total of $N = 6000$ feature samples.

In Fig. 4.3, we show the first level dependence structure (first tree of the R-Vine copula model; see Fig. 1.1 as an example) of the extracted features using our proposed R-Vine copula based fusion method for activity 'Walking-upstairs' in **ANGUITA** dataset. Here, features $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{h}_4$ are from the accelerometer while the features $\mathbf{h}_5, \mathbf{h}_6, \mathbf{h}_7$ are from the gyroscope. As we can see that, features $\mathbf{h}_3$ and $\mathbf{h}_6$ are highly dependent and the two modalities accelerometer and gyroscope are dependent mainly via these two features. Using the knowledge of intra-modal and cross-modal feature dependencies, we can trace back and find where these dependent features originated from, which would yield the reduction of training data needed in DNNs. Furthermore, we are able to understand the correlation among the raw data from different modalities. The R-Vine copula based fusion method adds interpretability of the model which explicitly provides the dependence structure for features from different modalities, compared to the neural network based fusion which is totally a 'black-box' model.

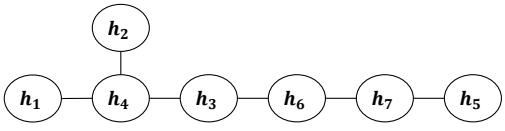

Fig. 4.3: First level dependence structure for activity 'Walking-upstairs'.

	Sit	Stand	Walk	Stairsup	Stairsdown	Bike	Recall
Sit	498	0	0	0	2	0	99.6%
Stand	0	454	0	0	46	0	90.8%
Walk	0	0	402	43	17	38	80.4%
Stairsup	0	0	24	408	50	0	81.6%
Stairsdown	0	0	39	53	408	0	81.6%
Bike	0	0	8	4	2	486	97.2%
Precision	100.0%	100.0%	85.0%	80.3%	77.7%	89.7%	88.6%

Table 4.3: **STISEN**: Confusion matrix for R-Vine copula based fusion.

	W	WU	WD	Si	St	L	Recall
Walking	276	0	17	3	0	0	93.2%
Walking-upstairs	6	259	0	3	3	0	95.6%.
Waking-downstairs	8	0	211	0	1	0	95.9%
Sitting	3	6	1	248	33	0	85.2%
Standing	3	3	2	26	298	0	87.4%
Laying	2	0	0	0	6	329	97.63%
Precision	92.6%	96.6%	91.3%	88.6%	87.4%	100.0%	92.8%

Table 4.4: **ANGUITA**: Confusion matrix for R-Vine copula based fusion.

Table 4.3 and Table 4.4 show the confusion matrices using the R-Vine copula based fusion scheme for the **STISEN** and **ANGUITA** datasets, respectively. As we observe, the fusion of phone and watch modalities achieves perfect classification for static activities ('Sit' and 'Stand'). Also, the fusion of the accelerometer and gyroscope from the smartphone achieves significantly accurate classification for moving activities (e.g., 'Walking', 'Laying').

4.5 Summary

In this chapter, an R-Vine copula based feature fusion approach was presented to perform activity recognition using multi-modal sensor observations. The features of each modality were extracted via a DNN and afterwards, an R-Vine copula model was constructed to capture the dependencies of intra-modal and cross-modal features. The procedures of model construction involve selecting the optimal R-Vine tree structure, obtaining the copula parameter set ϕ, and

choosing the best copula c^*. Experiments on two human activity datasets demonstrated the efficiency of our proposed method compared to neural network based data/feature fusion, in terms of high prediction accuracy, less number of training samples required and dependence interpretability.

CHAPTER 5
DISTRIBUTED SEQUENTIAL DETECTION WITH DEPENDENT OBSERVATIONS

5.1 Motivation

Distributed detection problems in sensor networks with *fixed-sample-size* (FSS) have been studied extensively [17, 28, 33, 54, 57, 58, 100, 105, 109], where the goal often is to minimize the probability of detection error at the FC based on a fixed number of observations collected by the sensors. For distributed detection problems in sensor networks, the challenge is to achieve high performance in terms of accuracy efficiency and time efficiency while satisfying energy and bandwidth constraints. In terms of detection performance, one critical issue for distributed detection problems in sensor networks is that the observations at the spatially distributed sensors may be highly dependent. Also, in terms of the detection time, sequential (*random-sample-size*) methods have shown their ability to improve time efficiency compared to FSS methods.

Wald's sequential probability ratio test (SPRT) [107] has been shown to be the optimal sequential test that arrives at a decision as soon as possible for binary hypothesis testing problems

by ignoring overshoots. However, the average detection time for SPRT can be larger than that of FSS tests under the same error probabilities for some cases [10, 97]. A truncated SPRT was proposed in [98], where the truncation time was chosen based on the corresponding FSS test and the average detection time always stayed below that of the FSS test at the expense of a small increase in error probabilities.

The problem of distributed sequential detection based on SPRT has attracted a lot of attention [34, 47, 65, 69, 77, 106]. In [34, 77, 106], likelihood-ratio-based quantizers were employed at the local sensors and a generalized SPRT (GSPRT) was used at the FC based on the quantized messages sent by the local sensors. In [47, 69], local tests and the test at the FC were both chosen to be GSPRT. In [65], a distributed sequential binary hypothesis testing scheme was proposed, where a GSPRT was performed at the FC and a level-triggered sampling scheme was proposed at the local sensors. However, in the aforementioned literature [34, 47, 65, 69, 77, 106], observations at the sensors were assumed to be independent and their spatial dependence was not considered.

In this chapter, we consider a distributed sequential detection problem with spatially dependent sensor observations. We assume that the channels from the sensors to the FC are corrupted by additive noise, including possibly non-Gaussian noise. We propose a distributed copula-based sequential scheme, where sequential tests are conducted at both the local sensors and the FC. More specifically, we perform a R-Vine copula based SPRT at the FC and memory-less grouped-data truncated SPRTs at the local sensors.

5.2 Problem Formulation

Consider a sequential binary hypothesis testing problem for the sensor network shown in Fig. 5.1. The two hypotheses, denoted by H_1 and H_0, are associated with the random phenomenon of interest that is monitored continuously by L sensors. Here, H_1 denotes the presence of the phenomenon and H_0 denotes the absence of the phenomenon. Suppose that the

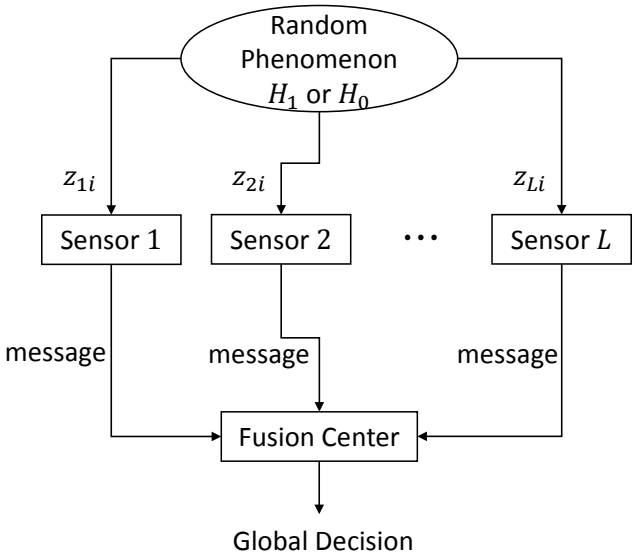

Fig. 5.1: Parallel distributed detection system.

l^{th} sensor acquires the observation $z_{li}, l = 1, 2, \ldots, L$ at each time instant $i = \{1, 2, \ldots\}$, and forwards its raw or compressed version of the data over a noisy channel to the FC that runs a sequential test and produces a global decision based on the received messages from the sensors. At the FC, the sequential procedure has three possible outcomes: it may either 1) accept H_0 and stop the test, or 2) accept H_1 and stop the test or 3) make no decision and acquire a new observation. The FC repeats this process until a decision is reached, in which case the test stops. Let T denote the stopping time. The goal is to minimize the expected stopping time $\mathbb{E}_k[T]$ under hypothesis $k = 0, 1$ given that $P_F \leq \alpha, P_M \leq \beta$, where P_F is the global probability of false alarm with constraint $\alpha \in (0, 1/2)$ and P_M is the global probability of miss detection with constraint $\beta \in (0, 1/2)$. We first make the following assumptions.

- Sensor observations $z_{li}, l = 1, 2, \ldots, L, i = 1, 2, \ldots$ are continuous random variables and i.i.d. over time.

- The PDFs of sensor observations $z_{li}, l = 1, 2, \ldots, L$, namely the marginal PDFs, are known under both hypotheses and given by $g_{k,l}(z_{li}), k = 0, 1$.

- The channel links between the sensors and the FC are corrupted by additive noise $w_{lj}, l = 1, 2, \ldots, L$ and j is the receiving time instant at the FC. Also, $w_{lj}, l = 1, 2, \ldots, L$ are assumed to be i.i.d. over time and independent of the messages sent by the local sensors.

- The signal received at the FC corresponding to sensor l at time j, after being corrupted by the imperfect channel, is denoted by y_{lj}.

- The marginal PDFs $f_{k,l}(y_{lj})$ and CDFs $F_l^k(y_{lj})$ with $k = 0, 1$ under both hypotheses are assumed to be known.

- The received observations $y_{1j}, \ldots, y_{Lj}$ at the FC and time instant j are assumed to be spatially dependent, namely, $f_k(\mathbf{y}_j) \neq \prod_{l=1}^{L} f_{k,l}(y_{lj}), k = 0, 1$, where $f_k(\cdot)$ and $f_{k,l}(\cdot)$ denote the joint PDF and marginal PDF for sensor l, conditioned on H_k, respectively. $\mathbf{y}_j = [y_{1j}, \ldots, y_{Lj}]$ is the observation vector for all the sensors at time instant j. We use the copula density function $c_k(\cdot|\boldsymbol{\phi}_k)$ to characterize the underlying dependence under hypothesis $H_k, k = 0, 1$ (see Equation (1.2)), which is typically not available *a priori*. Note that this dependence may result from the dependent messages sent by the local sensors, the dependent additive noise or both.

Remark 5.1. *In some applications, we may not have any prior information related to the phenomenon of interest, namely, the marginal PDFs and marginal CDFs may not be known. However, the marginal PDFs can be estimated non-parametrically using kernel density estimators [108] that provide a smoothed estimate of the true PDF by choosing the optimal bandwidth so that an accurate estimate is obtained. Moreover, the marginal CDF $F_l^k(\cdot)$ for sensor $l, l = 1, \ldots, L$ under hypothesis $k, k = 0, 1$ can be obtained by using Empirical Probability Integral Transforms in [45].*

Before we proceed, we recall the definition of Kullback-Leibler (KL) divergence between

two PDFs $f(x)$ and $g(x)$, denoted by $D(f(x)||g(x))$, given as

$$D(f(x)||g(x)) = \int f(x)\log\left(\frac{f(x)}{g(x)}\right)dx.$$

Moreover, for any PDFs $f(x)$ and $g(x)$, $D\left(f(x)||g(x)\right) \geq 0$ with equality if and only if $f(x) = g(x)$. Throughout this chapter, we assume that

A_1 There exists at least one sensor l, for which $D\left(f_{0,l}(\cdot)||f_{1,l}(\cdot)\right)$ and $D\left(f_{1,l}(\cdot)||f_{0,l}(\cdot)\right)$ are finite and positive. Moreover,

$$0 < D\left(c_0(\cdot|\boldsymbol{\phi}_0)||c_1(\cdot|\boldsymbol{\phi}_1)\right) < \infty,$$
$$0 < D\left(c_1(\cdot|\boldsymbol{\phi}_1)||c_0(\cdot|\boldsymbol{\phi}_0)\right) < \infty.$$

Remark 5.2. *Note that the condition A_1 guarantees that the two hypotheses are distinguishable, i.e., $f_1(\cdot)$ is not equal to $f_0(\cdot)$ almost everywhere, where $f_k(\cdot)$ denotes the joint PDF under hypothesis $H_k, k = 0, 1$.*

In the following sections, we start by designing a centralized copula-based SPRT; then propose a distributed copula-based SPRT with memory-less grouped-data SPRT tests performed at the local sensors.

Remark 5.3. *In this work, we ignore the overshoot effects while designing the sequential SPRTs in the following sections.*

5.3 Centralized Copula-based Sequential Probability Ratio Test

In this section, we consider the centralized sequential test. In this case, corrupted raw observations $y_{li} = z_{li} + w_{li}, l = 1, 2, \ldots, L, i = 1, 2, \ldots$ are received at the FC sequentially. The

FC performs a copula-based SPRT to obtain the final decision. More specifically, we solve the following binary hypothesis testing problem at the FC:

$$\begin{aligned} H_0 &: \mathbf{y}_i \sim f_0^{\text{cen}}(\mathbf{y}_i), i = 1, 2, \ldots, \\ H_1 &: \mathbf{y}_i \sim f_1^{\text{cen}}(\mathbf{y}_i), i = 1, 2, \ldots, \end{aligned} \tag{5.1}$$

where $\mathbf{y}_i = [y_{1i}, \ldots, y_{Li}]$, f_1^{cen} and f_0^{cen} denote the joint PDFs for the centralized case under alternative and null hypotheses, respectively. We take the existing dependence into account and design a copula-based SPRT at the FC.

Using Sklar's theorem, i.e., Theorem 1.1 in Section 1.1.1, the joint PDFs in Equation (5.1) for $i = 1, 2, \ldots, n$ can be expressed in terms of the marginal distributions and the copula densities, c_1^{cen} and c_0^{cen}, respectively, under H_1 and H_0, as

$$\begin{aligned} f_0^{\text{cen}}(\mathbf{y}) &= \prod_{i=1}^{n} \prod_{l=1}^{L} f_{0,l}^{\text{cen}}(y_{li}) \times c_0^{\text{cen}}(\mathbf{F}^{0,\text{cen}}(\mathbf{y}_i) | \boldsymbol{\phi}_0^{\text{cen}}), \\ f_1^{\text{cen}}(\mathbf{y}) &= \prod_{i=1}^{n} \prod_{l=1}^{L} f_{1,l}^{\text{cen}}(y_{li}) \times c_1^{\text{cen}}(\mathbf{F}^{1,\text{cen}}(\mathbf{y}_i) | \boldsymbol{\phi}_1^{\text{cen}}), \end{aligned} \tag{5.2}$$

where $f_{k,l}^{\text{cen}}(y_{li})$ are the marginal PDFs and $\mathbf{F}^{k,\text{cen}}(\mathbf{y}_i) = [F_1^{k,\text{cen}}(y_{1i}), \ldots, F_L^{k,\text{cen}}(y_{Li})]$ are the marginal CDFs for all the sensors at time instant i under hypothesis $k, k = 0, 1$. Moreover, $\boldsymbol{\phi}_0^{\text{cen}}$ and $\boldsymbol{\phi}_1^{\text{cen}}$ are the parameters of copula c_0^{cen} and c_1^{cen}, respectively.

For known $c_0^{\text{cen}}(\cdot|\boldsymbol{\phi}_0^{\text{cen}})$ and $c_1^{\text{cen}}(\cdot|\boldsymbol{\phi}_1^{\text{cen}})$, the centralized copula-based SPRT at the FC follows the following procedure: for $n = 1, 2, \ldots$,

$$\begin{cases} \Lambda^{n,\text{cen}}(\mathbf{y}) \geq A, & \text{decide } H_1, \\ \Lambda^{n,\text{cen}}(\mathbf{y}) \leq -B, & \text{decide } H_0, \\ -B < \Lambda^{n,\text{cen}}(\mathbf{y}) < A, & \text{take another observation}, \end{cases} \tag{5.3}$$

where A and $-B$ are the upper and lower thresholds, respectively, which are predetermined

constants such that $P_F \leq \alpha$ and $P_M \leq \beta$. Also, $\Lambda^{n,\text{cen}}(\mathbf{y})$ is given as

$$\begin{aligned}\Lambda^{n,\text{cen}}(\mathbf{y}) = &\sum_{i=1}^{n}\sum_{l=1}^{L}\log\frac{f_{1,l}^{\text{cen}}(y_{li})}{f_{0,l}^{\text{cen}}(y_{li})} \\ &+\sum_{i=1}^{n}\log\frac{c_1^{\text{cen}}(\mathbf{F}^{1,\text{cen}}(\mathbf{y}_i)|\boldsymbol{\phi}_1^{\text{cen}})}{c_0^{\text{cen}}(\mathbf{F}^{0,\text{cen}}(\mathbf{y}_i)|\boldsymbol{\phi}_0^{\text{cen}})}.\end{aligned} \tag{5.4}$$

In general, for given α and β, exact analytical expressions of the optimal thresholds A and $-B$ are intractable. One may use the approximated thresholds obtained from Wald's SPRT [107], which are given by

$$A \approx \log\frac{1-\beta}{\alpha}, \quad -B \approx \log\frac{\beta}{1-\alpha}, \tag{5.5}$$

where if $\alpha, \beta \in (0, \frac{1}{2})$, we have $-B < A$.

Let N be the stopping time for the centralized scheme. Since the messages received at the FC are i.i.d., we have $P(N < \infty|H_k) = 1, k = 0, 1$ under condition A_1 [99, Lemma 3.1.1]. The goal for the above centralized copula-based SPRT is to minimize the average stopping time $\mathbb{E}_k[N], k = 0, 1$ such that $P_F \leq \alpha, P_M \leq \beta$. In Theorem 5.1, we show the asymptotic optimality of the centralized copula-based SPRT as $A, B \to \infty$. Note that in this work, we assume that when the test stops, we have $\Lambda^{n,\text{cen}}(\mathbf{y}) = A$ or $\Lambda^{n,\text{cen}}(\mathbf{y}) = -B$ by ignoring the overshoots. Therefore, we use $\approx$ while evaluating P_F, P_M and the expected stopping time $\mathbb{E}_{H_k}[N], k = 0, 1$ in Theorems 5.1 and 5.3.

Theorem 5.1. *For the centralized copula-based SPRT in Equation* (5.3)*, as $A \to \infty$ and $B \to \infty$, we have $P_F \approx e^{-A}$ and $P_M \approx e^{-B}$. Moreover, the average stopping times under the two hypotheses admit the following asymptotic expressions:*

$$\mathbb{E}_{H_0}[N] \approx \frac{B}{D_0^{cen}}, \quad \mathbb{E}_{H_1}[N] \approx \frac{A}{D_1^{cen}}, \tag{5.6}$$

as $A, B \to \infty$, and where

$$D_0^{cen} = \sum_{l=1}^{L} D\left(f_{0,l}^{cen}(\cdot)||f_{1,l}^{cen}(\cdot)\right) + D\left(c_0^{cen}(\cdot|\boldsymbol{\phi}_0^{cen})||c_1^{cen}(\cdot|\boldsymbol{\phi}_1^{cen})\right),$$
$$D_1^{cen} = \sum_{l=1}^{L} D\left(f_{1,l}^{cen}(\cdot)||f_{0,l}^{cen}(\cdot)\right) + D\left(c_1^{cen}(\cdot|\boldsymbol{\phi}_1^{cen})||c_0^{cen}(\cdot|\boldsymbol{\phi}_0^{cen})\right).$$

Proof: See Appendix C. ■

Remark 5.4. *The asymptotic performance obtained in Equation* (5.6) *shows that the average detection time depends on the KL distance provided by each sensor as well as the KL distance due to the spatial dependence among L sensors. By including the spatial dependence in our analysis, we can reduce the detection time on an average.*

Typically, the copula density function $c_k^{\text{cen}}(\cdot|\boldsymbol{\phi}_k^{\text{cen}})$ and its corresponding parameter set $\boldsymbol{\phi}_k^{\text{cen}}$ under hypothesis $H_k, k = 0, 1$ are not known and need to be estimated. Using maximum likelihood estimates in place of the true copula density functions and the true parameters, the centralized copula-based SPRT in Equation (5.3) becomes a generalized copula-based SPRT. In the following, we present the estimation of the best copula model.

Since the FC has no knowledge of the dependence structure of the received messages, we assume that the FC waits for N_0 messages before starting the copula-based SPRT. Note that N_0 can determined by the goodness-of-fit tests for copula models [36]. Hence, the copula density functions and their corresponding parameters can be estimated. The estimation of optimal copula density functions under the two hypotheses is similar. Therefore, the hypothesis index k will be omitted for now to simplify notations. Also, the superscript “cen" for the centralized sequential test is omitted for now. Note that the marginal CDFs at each time instant i are needed to evaluate the copula log likelihood ratios as shown in Equation (5.4).

We apply R-Vine copula to estimate the multivariate copula function $c_k^{\text{cen}}(\cdot|\boldsymbol{\phi}_k^{\text{cen}})$ under hypothesis $H_k, k = 0, 1$. The estimation of R-Vine copula model requires the determination of

the R-Vine tree structure, the choice of bivariate copulas and the estimation of their corresponding parameters. To select the optimal R-Vine tree structure, we adopt the sequential maximum spanning tree algorithm in [27]. This sequential method is based on Kendall's τ. The sample estimate of Kendall's τ, for N_0 observations, can be calculated using Equation (1.10).

The estimation of the R-Vine copula model is presented in Algorithm 2.1 with $m = L$ and data sequence $(y_{1j}, \ldots, y_{Lj}), j = 1, 2, \ldots, N_0$. The main computational complexity of this algorithm is from the fit of the bivariate copulas. For a d-dimensional system, the number of bivariate copulas needed to fit in the tree is $d(d-1)/2$. Therefore, the computational complexity is $\mathcal{O}(d^2)$. In the following, we describe the sequential estimation in detail. The sequential method starts with the selection of the first tree T_1 and continues tree by tree up to the last tree T_{L-1}. The trees are selected in a way that the chosen bivariate copula models the strongest pair-wise dependencies present which are characterized by Kendall's τ.

After the R-Vine tree structure is obtained, we need to define a bivariate copula set and estimate the optimal bivariate copulas that best characterize the pair-wise dependencies. Consider a library of copulas, $\mathcal{C} = \{c_m : m = 1, \ldots, M\}$. Before estimating the optimal bivariate copula, the copula parameter set $\boldsymbol{\phi}$ is first obtained using MLE as

$$\hat{\boldsymbol{\phi}} = \arg\max_{\boldsymbol{\phi}} \sum_{i=1}^{N_0} \log c \left(\hat{F}_{l_1}(y_{l_1 i}), \hat{F}_{l_2}(y_{l_2 i}) | \boldsymbol{\phi} \right), \tag{5.7}$$

where $(l_1, l_2), l_1, l_2 \in [L]$ is a connected pair in the selected R-Vine trees and for simplification of notation, we omit the conditioning elements for conditional marginal CDFs. Note that the conditional marginal CDFs can be obtained recursively using Equation (1.12).

Once the copula parameter set is obtained, the best copula c^* is selected from the copula library $\mathcal{C}$ using the Akaike Information Criterion (AIC) [3] as the criterion, given as

$$\text{AIC}_m = -\sum_{i=1}^{N_0} \log c_m \left(\hat{F}_{l_1}(y_{l_1 i}), \hat{F}_{l_2}(y_{l_2 i}) | \hat{\boldsymbol{\phi}}_m \right) + 2q(m), \tag{5.8}$$

where $q(m)$ is the number of parameters in the mth copula model. Also, the conditioning elements for conditional marginal CDFs are omitted.

The best copula c^* is

$$c^* = \arg\min_{c_m \in \mathcal{C}} \text{AIC}_m. \tag{5.9}$$

Now, we have the optimal copula density functions $c_1^{\text{cen},*}(\cdot|\hat{\boldsymbol{\phi}}_1^{\text{cen},*})$ and $c_0^{\text{cen},*}(\cdot|\hat{\boldsymbol{\phi}}_0^{\text{cen},*})$ under the alternative and null hypotheses, respectively. The log-likelihood ratio test statistics $\Lambda^{n,\text{cen}}(\mathbf{y})$ in Equation (5.4) now becomes

$$\begin{aligned}\Lambda^{n,\text{cen}}(\mathbf{y}) = &\sum_{i=1}^{n}\sum_{l=1}^{L} \log \frac{f_{1,l}^{\text{cen}}(y_{li})}{f_{0,l}^{\text{cen}}(y_{li})} \\ &+ \sum_{i=1}^{n} \log \frac{c_1^{\text{cen},*}(\mathbf{F}^{1,\text{cen}}(\mathbf{y}_i)|\hat{\boldsymbol{\phi}}_1^{\text{cen},*})}{c_0^{\text{cen},*}(\mathbf{F}^{0,\text{cen}}(\mathbf{y}_i)|\hat{\boldsymbol{\phi}}_0^{\text{cen},*})}.\end{aligned} \tag{5.10}$$

Note that the centralized copula-based SPRT incurs substantial data transmission overhead between the sensors and the FC. In the following section, we proceed to develop the distributed copula-based SPRT.

5.4 Distributed Copula-based Sequential Probability Ratio Test

In this section, we propose a distributed copula-based SPRT, where the FC receives noise corrupted binary messages instead of the raw measurements. In particular, the local sensors perform truncated SPRT tests and send binary decisions to the FC. In the following, we first present the local sensor detection rule and then obtain the FC fusion rule.

5.4.1 Local Sensor Detection Rule

To design the local sensor detection rule, we need to determine the communication protocol and binary messages to be sent to the FC at each sensor. Here, we propose a memory-less grouped-data sequential test to be carried out at each sensor. Consider the following sequential procedure: at each stage, each sensor sequentially acquires at most W_0 observations over the observation window, makes a binary decision by performing a local SPRT, and transmits the decision to the FC. After transmission, each sensor moves to the next stage of the test, refreshes its memory and runs another SPRT based on newly arriving measurements. Therefore, the FC receives i.i.d. messages over time. Note that there are two cases for the local SPRTs: Case 1) sensors make the decision before or at the time that they take the W_0^{th} measurement, Case 2) sensors do not make any decision after taking the W_0^{th} measurement. In Case 1), we let the sensors go to sleep mode after each transmission. After the FC receives the messages from all the sensors, it sends a wakeup signal to the local sensors, and sensors wake up and start the next stage of the test. In Case 2), since a decision has not been made, we truncate the test after taking the W_0^{th} measurement. The group size W_0 is also referred to as the truncation window at all the sensors [60]. Thus, all the local sensors transmit their decisions at $t_l, l = 1, \ldots, L$, where t^l is the stopping time for sensor l and $t^l \leq W_0, \forall l$. Let $t_{\max} = \max\{t^1, \ldots, t^L\}$ be the maximum stopping time among all the sensors. The next stage of the proposed local tests started at $t_{\max}$, the time that the FC sends the wakeup signal.

The memory-less grouped-data local SPRT is defined as follows. At stage i, sensors perform the local SPRTs based on $(i-1)W_0 + 1^{th}$ to $iW_0{}^{th}$ measurements. The local log-likelihood ratio test statistics $\lambda_{l,i}(W), l = 1, \ldots, L$, are given as follows. For $(i-1)W_0 + 1 \leq W \leq iW_0$,

$$\lambda_{l,i}(W) = \sum_{j=(i-1)W_0+1}^{W} \log \frac{g_{1,l}(z_{lj})}{g_{0,l}(z_{lj})}. \tag{5.11}$$

The binary messages U_{li}^W are given as:

$$\begin{cases} \text{for } (i-1)W_0+1 \le W \le iW_0, \\ U_{li}^W = \begin{cases} 1, & \text{if} \quad \lambda_{l,i}(W) \ge \quad a_l, \\ 0, & \text{if} \quad \lambda_{l,i}(W) \le -b_l, \\ \infty, & \text{if} \quad -b_l < \lambda_{l,i}(W) < a_l, \end{cases} \\ \text{for } W = iW_0 \text{ and } U_{li}^W = \infty, \\ U_{li}^W = \begin{cases} 1, & \text{if} \quad \lambda_{l,i}(W) \ge \nu_l, \\ 0, & \text{if} \quad \lambda_{l,i}(W) < \nu_l, \end{cases} \end{cases} \tag{5.12}$$

where a_l, $-b_l$ and ν_l are the thresholds with $-b_l \le \nu_l \le a_l$. By moving the thresholds a_l and $-b_l$ away from each other, we can decrease P_F^{local} and P_M^{local} with an increased average stopping time. There is a tradeoff between the choice of the thresholds and the truncation time. Therefore, our goal here is to design the thresholds jointly so that the local probability of false alarm P_F^{local} and the local probability of miss detection P_M^{local} stay below pre-specified constraints on error probabilities. In the following, we first design the thresholds a_l and $-b_l$ based on the untruncated SPRT, and then design the threshold ν_l and the truncation window W_0 using the corresponding FSS test so that in the worst case, the number of samples at local detectors is equal that of the FSS test.

Now, we design the thresholds a_l and $-b_l$ based on the untruncated SPRT, where the constraints of the probabilities of false alarm and miss detection are $\tilde{\alpha}$ and $\tilde{\beta}$, respectively. Let $Q_k(x) = P(\lambda_{l,i}(W) \le x | H_k)$. We require that $Q_0(a_l) = 1 - \tilde{\alpha}/\Delta_{l,a}$ and $Q_1(-b_l) = \tilde{\beta}/\Delta_{l,b}$, where $\Delta_{l,a}, \Delta_{l,b} \ge 1$ are the scale parameters. Note that we may not have closed forms of $Q_1(\cdot)$ and $Q_0(\cdot)$. By choosing appropriate scale parameters $\Delta_{l,a}$ and $\Delta_{l,b}$, similar to the thresholds of centralized copula-based sequential test in Equation (5.5), we can obtain the approximated thresholds as

$$a_l \approx \log \frac{1-\tilde{\beta}/\Delta_{l,b}}{\tilde{\alpha}/\Delta_{l,a}}, \quad -b_l \approx \log \frac{\tilde{\beta}/\Delta_{l,b}}{1-\tilde{\alpha}/\Delta_{l,a}}. \tag{5.13}$$

Before we design the threshold ν_l and the truncation window W_0, we first present the corresponding FSS test for sensor $l, l = 1, \ldots, L$, which has the same constraints of error probabilities as the untruncated SPRT.

$$U_{li}^{FSS} = \begin{cases} 1, & \text{if} \quad \lambda_{l,i}^{FSS} \geq \quad \nu_l^{FSS}, \\ 0, & \text{if} \quad \lambda_{l,i}^{FSS} < \nu_l^{FSS}, \end{cases} \tag{5.14}$$

where $\lambda_{l,i}^{FSS} = \sum_{j=(i-1)N_{FSS}+1}^{iN_{FSS}} \log \frac{g_{1,l}(z_{lj})}{g_{0,l}(z_{lj})}$ and N_{FSS} is the number of observations needed for the FSS test. ν_l^{FSS} is the threshold. Also, N_{FSS} and ν_l^{FSS} are designed so that the local probabilities of false alarm and miss detection stay below levels $\tilde{\alpha}/\Delta_{l,Ta}$ and $\tilde{\beta}/\Delta_{l,Tb}$, where $\Delta_{l,Ta}, \Delta_{l,Tb} \geq 1$ are the scale parameters to be designed.

Remark 5.5. *The scale parameters $\Delta_{l,a}, \Delta_{l,b}, \Delta_{l,Ta}$ and $\Delta_{l,Tb}$ are introduced to adjust the levels of the local probability of false alarm and the local probability of miss detection so that P_F^{local} and P_M^{local} are upper bounded by certain pre-specified error probabilities.*

In the following Theorem, we show that by setting W_0 equal to N_{FSS} and ν_l equal to ν_l^{FSS}, we can obtain the upper bounds of P_F^{local} and P_M^{local}.

Theorem 5.2. *The upper bounds of P_F^{local} and P_M^{local} are given as:*

$$P_F^{local} < \left(\frac{1}{\Delta_{l,a}} + \frac{1}{\Delta_{l,Ta}} \right) \tilde{\alpha}, \tag{5.15}$$

$$P_M^{local} < \left(\frac{1}{\Delta_{l,b}} + \frac{1}{\Delta_{l,Tb}} \right) \tilde{\beta}. \tag{5.16}$$

Proof: See Appendix D. ■

Remark 5.6. *According to Theorem 5.2, P_F^{local} and P_M^{local} can be determined by choosing appropriate $\Delta_{l,a}$, $\Delta_{l,b}$, $\Delta_{l,Ta}$ and $\Delta_{l,Tb}$. Since $P(U_{li}^T = \infty | H_k) < 1, k = 0, 1, (i-1)W_0 + 1 \leq W \leq iW_0$, we have $\mathbb{E}[t^l] < W_0$. Compared to the FSS test, the local sensors need fewer observations on average.*

5.4.2 Derivation of the Fusion Rule

We omit the index W from the binary decision U_{li}^{W} and assume that the FC receives $U_{1i}, \ldots, U_{Li}$ sequentially from all the sensors. After receiving the noise corrupted local sensor decisions $y_{ln} = U_{ln} + w_{ln}, l = 1, \ldots, L$, where n is the receiving time, the FC combines all received local sensor messages based on copula theory. Therefore, the FC performs the following global copula-based SPRT,

$$\begin{cases} \Lambda^n(\mathbf{y}) \geq A, & \text{decide } H_1, \\ \Lambda^n(\mathbf{y}) \leq -B, & \text{decide } H_0, \\ -B < \Lambda^n(\mathbf{y}) < A, & \text{take another observation}, \end{cases} \tag{5.17}$$

where A and $-B$ are the upper and lower thresholds, respectively, which are determined by the global probability of false alarm α and global probability of miss detection β. Also, $\Lambda^n(\mathbf{y})$ is given as

$$\Lambda^n(\mathbf{y}) = \sum_{j=1}^{N_n} \sum_{l=1}^{L} \log \frac{f_{1,l}(y_{lj})}{f_{0,l}(y_{lj})} + \sum_{j=1}^{N_n} \log \frac{c_1(\mathbf{F}^1(\mathbf{y}_j)|\boldsymbol{\phi}_1)}{c_0(\mathbf{F}^0(\mathbf{y}_j)|\boldsymbol{\phi}_0)}, \tag{5.18}$$

where $\mathbf{F}^k(\mathbf{y}_j) = [F_1^k(y_{1j}), \ldots, F_L^k(y_{Lj})]$, N_n is the number of messages received by the FC at time n and $c_k(\cdot|\boldsymbol{\phi}_k)$ is the copula density function with its corresponding parameter $\boldsymbol{\phi}_k$ under hypothesis $k, k = 0, 1$. Let T_p denote the stopping time at the FC.

Remark 5.7. *Note that the marginal PDFs of y_{lj} under H_1 and H_0 are given as*

$$f_{1,l}(y_{lj}) = (1 - P_M^{local}) f_w^1(1 + w_{lj}) + P_M^{local} f_w^1(w_{lj}),$$
$$f_{0,l}(y_{lj}) = (1 - P_F^{local}) f_w^0(w_{lj}) + P_F^{local} f_w^0(1 + w_{lj}),$$

where $f_w^k(\cdot)$ is the marginal PDF of the channel noise from local sensors to the FC under $H_k, k = 0, 1$. Also, by assuming the $\lambda_{l,i}(W), (i-1)W_0 + 1 \leq W \leq iW_0$ to be Gaussian distributed, the P_M^{local} and P_F^{local} can be computed analytically [112]. Without the Gaussian

assumption, the P_M^{local} and P_F^{local} can be obtained empirically.

Assuming that the condition A_1 is satisfied, the asymptotic performance of the distributed copula-based SPRT is given by the following Theorem.

Theorem 5.3. *In the asymptotic regime where $A, B \to \infty$, we have*

$$P_F \approx e^{-A}, \quad P_M \approx e^{-B}, \tag{5.19}$$

and in the worst case, the minimum average stopping times of the distributed copula-based SPRT in Equation (5.17) *under the two hypotheses have the following asymptotic performance:*

$$\mathbb{E}_{H_0}[T_p] \approx \frac{B * W_0}{D_0}, \tag{5.20}$$

and

$$\mathbb{E}_{H_1}[T_p] \approx \frac{A * W_0}{D_1}, \tag{5.21}$$

where $D_0 = \sum_{l=1}^{L} D\left(f_{0,l}(\cdot)||f_{1,l}(\cdot)\right) + D\left(c_0(\cdot|\boldsymbol{\phi}_0)||c_1(\cdot|\boldsymbol{\phi}_1)\right)$ and $D_1 = \sum_{l=1}^{L} D\left(f_{1,l}(\cdot)||f_{0,l}(\cdot)\right) + D\left(c_1(\cdot|\boldsymbol{\phi}_1)||c_0(\cdot|\boldsymbol{\phi}_0)\right)$.

Proof: See Appendix E. ■

Remark 5.8. *Compared to the distributed SPRT that ignores the underlying dependence, our proposed distributed copula-based SPRT is more efficient on an average in terms of the detection time. Moreover, since the FC may not have any knowledge of the underlying dependence, the copula density function c_k and its corresponding parameter $\boldsymbol{\phi}_k$ need to be estimated first. Similar to the centralized copula-based SPRT, the FC can estimate the unknown dependence using the R-Vine copula based methods using a methodology similar to the centralized copula-based SPRT.*

Since quantization cannot increase the value of KL divergence [103], we have $D_k/W_0 \leq D_k^{\text{cen}}$ under hypothesis $k, k = 0, 1$. The performance of the distributed copula-based SPRT

degrades compared to the centralized copula-based SPRT. Moreover, the performance also depends on the local thresholds a_l, $-b_l$ and ν_l for $l = 1, 2, \ldots, L$. These thresholds need to be chosen such that D_0 and D_1 are maximized. However, in general, it is not feasible to find the optimal local thresholds due to the existing complex dependence and the channel noise.

5.5 Simulation Results

In this section, we demonstrate the efficacy of our proposed copula-based SPRTs through numerical examples. We assume that there are two hypotheses, where H_1 denotes the presence of a signal s and H_0 indicates the absence of s. Also, s is assumed to be a deterministic signal. We model the signals received at the sensors as:

$$\begin{aligned} H_1 &: z_{li} = s + v_{li}, \qquad && l = 1, \ldots L; i = 1, 2, \ldots, \\ H_0 &: z_{li} = v_{li}, && l = 1, \ldots, L; i = 1, 2, \ldots, \end{aligned} \tag{5.22}$$

where v_{li} denotes the measurement noise at sensor l and time instant i. We assume that v_{li} follows a zero-mean Gaussian distribution with standard deviation $\sigma_v^l, l = 1, \ldots L$. The received signal z_{li} is assumed to be temporally independent conditioned on either hypothesis but can be spatially dependent.

The FC receives $y_{ln} = x_{ln} + w_{ln}$, where n is the receiving time instant at the FC and x_{ln} is the message sent by the l^{th} sensor. For distributed detection, we have $x_{ln} = U_{ln}$. Similarly, $x_{ln} = z_{ln}$ for centralized detection. Moreover, w_{ln} is the channel noise from the sensors to the FC and drawn from a zero-mean Gaussian distribution with standard deviation $\sigma_w^l, l = 1, \ldots L$. Moreover, the channel noise is assumed to be temporally independent conditioned on either hypothesis but can be spatially dependent. Also, the channel noise and the signals at the local sensors are assumed to be mutually independent.

Unless specified otherwise, we assume that $\sigma_w^l = \sqrt{3}$ and $\sigma_v^l = 1$. We use $N_0 = 100$

observations to estimate the R-Vine copulas. For each sensor, the local probability of false alarm and miss detection constraints are $\tilde{\alpha} = 0.01$ and $\tilde{\beta} = 0.01$, respectively. Also, the scale parameters are set as $\Delta_{l,a} = \Delta_{l,b} = \Delta_{l,Ta} = \Delta_{l,Tb} = 1$ for $l = 1, 2, \ldots, L$. Hence, the upper bounds of local probability false alarm and miss detection for each sensor are $P_F^{\text{local}} \leq 2\tilde{\alpha}$ and $P_M^{\text{local}} \leq 2\tilde{\beta}$, respectively. Moreover, the global probability of false alarm and global miss detection constraints for all the sensors are $\alpha = 0.01$ and $\beta = 0.01$, respectively. Without loss of generality, the spatial dependence of sensor observations is generated using multivariate Gaussian copula except for Table 5.3. To exhibit the performance improvement by applying our proposed copula-based SPRTs, we also evaluate the performance of product-based SPRT that ignores dependence of sensor observations. For clarity, we summarize the empirically studied cases as follows.

- Case 1: We assume that the spatial dependence is resulting from target signals. The channel noises are assumed to be spatially and temporally independent.
- Case 2: We assume that spatial dependence is resulting from both the target signals and the channel noises.

Remark 5.9. *Without loss of generality, if the spatial dependence is resulting from target signals, we assume that under H_1, the observations $z_{1i}, \ldots, z_{Li}$ are dependent while under H_0, they are independent.*

Before we proceed to the results, we note that the expected stopping time $\mathbb{E}[T]$ in the following is defined as the average expected stopping time under hypothesis H_1 and H_0, namely, $\mathbb{E}[T] = \frac{1}{2}(\mathbb{E}_{H_1}[T] + \mathbb{E}_{H_0}[T])$. Also, $\mathbb{E}[T]$ is measured using a system-wide clock that is employed both at the local sensors and the FC.

In Table 5.1 and Table 5.2, we present the average P_F and P_M values as a function of α and β, respectively, where we compare the centralized product-based scheme and the centralized copula-based scheme for Case 1 with $L = 3$ for different signal-to-noise ratios (SNRs). The

	SNR $= -6$ dB				SNR $= -9$ dB			
	Centralized product based SPRT		Centralized copula based SPRT		Centralized product based SPRT		Centralized copula based SPRT	
β	P_F	P_M	P_F	P_M	P_F	P_M	P_F	P_M
0.4	0.0059	0.4812	0.0065	0.0433	0.0083	0.5693	0.0061	0.0408
0.3	0.0078	0.4341	0.0052	0.0321	0.0089	0.5106	0.0052	0.0299
0.2	0.0075	0.3786	0.0039	0.0186	0.0080	0.4366	0.0046	0.0205
0.1	0.0056	0.2849	0.0036	0.0101	0.0079	0.3362	0.0039	0.0099
0.01	0.0061	0.1245	0.0030	0.0010	0.0077	0.1416	0.0037	8.8000e-4
0.001	0.0060	0.0548	0.0032	1.1500e-04	0.0076	0.0617	0.0037	8.4000e-5
0.0001	0.0061	0.0240	0.0033	1.300e-05	0.0077	0.0270	0.0037	7.000e-6

Table 5.1: Known copula: Estimated P_F and P_M with $\alpha = 0.01$, $L = 3$ for centralized sequential scheme and Case 1.

	SNR $= -6$ dB				SNR $= -9$ dB			
	Centralized product based SPRT		Centralized copula based SPRT		Centralized product based SPRT		Centralized copula based SPRT	
α	P_F	P_M	P_F	P_M	P_F	P_M	P_F	P_M
0.4	0.2400	0.0903	0.1358	0.0013	0.3105	0.0842	0.1493	0.0016
0.3	0.1803	0.0918	0.1093	0.0018	0.2337	0.0914	0.1207	0.0010
0.2	0.1178	0.0968	0.0702	6.000e-4	0.1568	0.1027	0.0825	0.0011
0.1	0.0573	0.1019	0.0366	0.0010	0.0771	0.1151	0.0368	0.0012
0.01	0.0061	0.1245	0.0030	0.0010	0.0077	0.1416	0.0037	8.8000e-4
0.001	5.5100e-04	0.1352	3.1100e-4	9.7100e-04	8.2200e-4	0.1563	3.6900e-4	8.3000e-3
0.0001	4.7000e-5	0.1400	2.3000e-5	9.800e-04	7.400e-5	0.1617	3.3000e-5	9.8500e-4

Table 5.2: Known copula: Estimated P_F and P_M with $\beta = 0.01$, $L = 3$ for centralized sequential scheme and Case 1.

signal spatial dependence is assumed to be known. Since the underlying dependence is resulting from target signals, it implies that the sensor observations are independent under H_0. As we can see, the average P_M values obtained for the centralized copula-based SPRT satisfy the constraint β while those for the centralized product-based SPRT do not. The average P_F values are below α for both the centralized copula-based SPRT and the centralized product-based SPRT. This is because, sensor observations are independent under H_0. Also, in Fig. 5.2 and Fig. 5.3, we show the corresponding average expected stopping time $\mathbb{E}[T]$ varying α and β, respectively. As we observe, on average, the centralized copula-based SPRT makes decisions faster

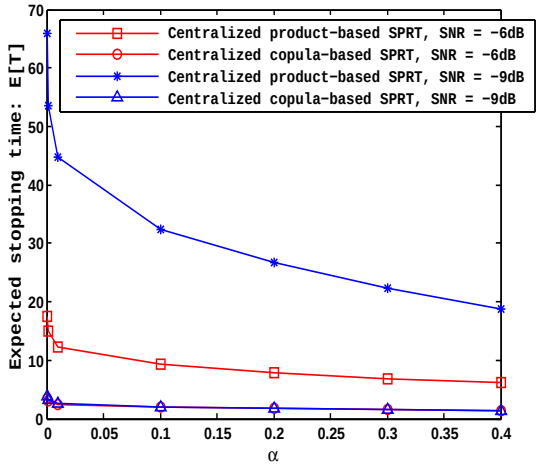

Fig. 5.2: Average expected stopping time as a function of α for centralized sequential scheme.

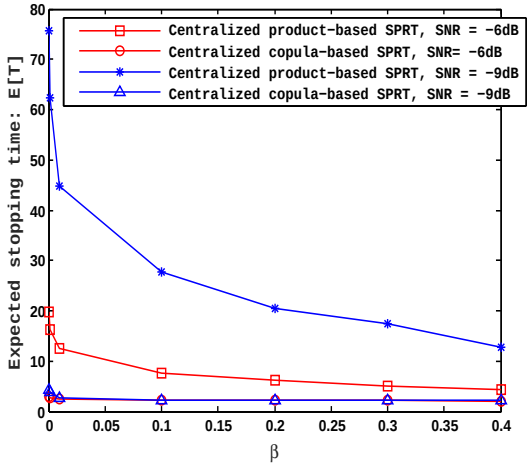

Fig. 5.3: Average expected stopping time as a function of β for centralized sequential scheme.

than the centralized product-based SPRT. Moreover, for lower SNRs, the centralized product-based SPRT requires more time to complete the detection while the centralized copula-based SPRT is less sensitive to low SNRs, i.e., does not change much for different values of SNR.

	Dependence generated using Gaussian copula		Dependence generated using R-Vine copula	
N_0	$L=3$	$L=5$	$L=3$	$L=5$
40	0.938	0.999	0.923	0.994
50	0.953	0.998	0.918	0.992
60	0.926	0.999	0.907	0.991
70	0.939	0.999	0.907	0.968
80	0.939	0.998	0.938	0.980
90	0.945	0.999	0.915	0.956
100	0.941	0.998	0.919	0.965
120	0.937	0.996	0.923	0.957
150	0.938	0.994	0.928	0.933
200	0.934	0.998	0.915	0.926

Table 5.3: Average p values for the estimation of underlying dependence using R-Vine copula model with different number of sensors.

Typically, we have no knowledge of the dependence information, especially for the distributed detection scheme. If the signals at the local sensors are dependent, then after local

	Dependence generated using R-Vine copula	
N_0	Clayton copula	Gaussian copula
40	0.037	0.143
50	0.016	0.026
60	0.020	0.019
70	0.022	0.029
80	0.001	0.030
90	0.002	0.015
100	0.002	0.049
120	0.000	0.003
150	0.000	0.002
200	0.000	0.001

Table 5.4: Average p values for the estimation of underlying dependence using different multivariate copula models with $L = 3$.

quantization this dependence structure may change. Therefore, we need to estimate the dependence structure, namely, the R-Vine copula. In Table 5.3, we present the average p values corresponding to the estimation of the underlying dependence using R-Vine copula models with different N_0s and number of sensors. We generated two types of dependence: linear dependence using multivariate Gaussian copula and non-linear dependence using R-Vine copula. Similar results can be obtained for the dependence generated by other multivariate copulas or even more complex dependence. As we can see, the R-Vine copula model performs very well and can not be rejected a 5% significance level. Note that the number of samples N_0 needed also depends on the complexity of the underlying dependence as well as the number of sensors.

To show that the R-Vine copula model is more capable of characterizing complex dependence compared to multivariate copula models, in Table 5.4, we show average p values for the estimation of the underlying dependence using Gaussian copula and Clayton copula models for $L = 3$. Here, the underlying dependence is generated using the R-Vine copula model. As we can see, for most of the cases, the Gaussian and Clayton copula models are rejected at a 5% significance level except when $N_0 = 40$, i. e., the p value is larger than .05 for the Gaussian

copula for $N_0 = 40$. However, the R-Vine copula model in Table 5.3 cannot be rejected at a 5% significance level, i. e., the p values are always greater than 0.05.

Remark 5.10. *Note that the p value is defined as the probability of the null hypothesis being true. In this case, we assume that the copula model C is unknown but belongs to a class $\mathcal{C}_0 = \{C_{\mathbf{\Phi}} : \mathbf{\Phi} \in \mathbb{R}^m\}$. We define the null hypothesis as $H_0 : \hat{C} \in \mathcal{C}_0$ and the alternative hypothesis as $H_1 : \hat{C} \notin \mathcal{C}_0$, where $\hat{C}$ is an estimate of C. More specifically, the hypothesis H_0 means $\hat{C}$ represents the distribution of C quite well. If the p value is in $[0, 0.05]$, the evidence is strong that one should reject the null hypothesis.*

In Fig. 5.4, we show the truncation window W_0 as a function of $\tilde{\alpha}$ and $\tilde{\beta}$ with $L = 3$. We can see that as $\tilde{\alpha}$ and $\tilde{\beta}$ decrease, the W_0 increases. Also, for lower SNR, more detection time is needed to achieve the same probabilities of error.

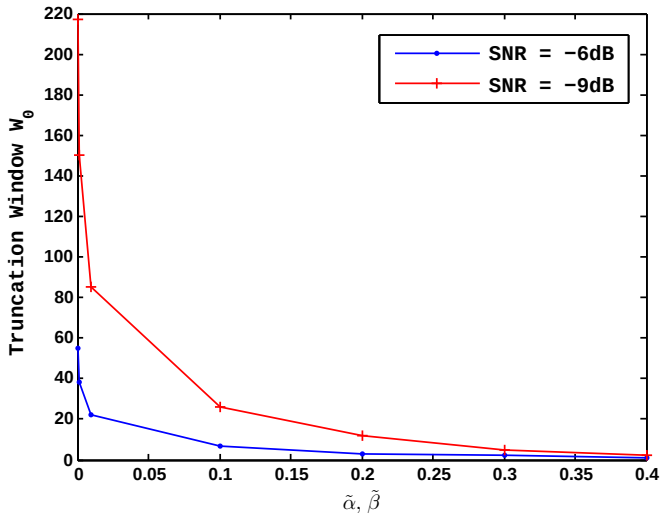

Fig. 5.4: Truncation window W_0 as a function of $\tilde{\alpha}$ and $\tilde{\beta}$, where $\tilde{\alpha} = \tilde{\beta}$.

In Table 5.5 and Table 5.6, we present the average P_F, P_M values and the average expected stopping time $\mathbb{E}[T]$ with unknown copulas by comparing the product-based SPRTs and the copula-based SPRTs (including the centralized and the distributed schemes) for the both cases with $L = 3$ under SNR $= -6$ dB and SNR $= -9$ dB, respectively. As we can see, for Case 1

	Case 1			Case 2		
	P_F	P_M	$\mathbb{E}[T]$	P_F	P_M	$\mathbb{E}[T]$
Centralized product based SPRT	0.0061	0.1254	12.518	0.0246	0.1254	12.317
Centralized copula based SPRT	0.0063	0.0038	2.625	0.0093	0.0012	2.847
Distributed product based SPRT	0.0048	0.0051	165.908	0.0215	0.1179	151.004
Distributed copula based SPRT	0.0053	0.0050	166.793	0.0072	0.0036	43.432

Table 5.5: Unknown copula: Estimated P_F, P_M and $\mathbb{E}[T]$ with $\alpha = \beta = 0.01$ and SNR $= -6$ dB.

	Case 1			Case 2		
	P_F	P_M	$\mathbb{E}[T]$	P_F	P_M	$\mathbb{E}[T]$
Centralized product based SPRT	0.0074	0.1427	44.802	0.0303	0.1421	43.937
Centralized copula based SPRT	0.0045	0.0007	2.645	0.0070	0.0010	3.100
Distributed product based SPRT	0.0050	0.0051	625.244	0.0215	0.1200	569.273
Distributed copula based SPRT	0.0054	0.0050	624.869	0.0093	0.0045	152.713

Table 5.6: Unknown copula: Estimated P_F, P_M and $\mathbb{E}[T]$ with $\alpha = \beta = 0.01$ and SNR $= -9$ dB.

and Case 2, the average P_F and P_M values satisfy the specified α and β values, respectively, for the copula-based SPRTs, while for the product-based SPRTs, the average P_F and P_M values do not satisfy the P_F and P_M constraints, except for the distributed product-based SPRT of Case 1. This is because, by setting the scale parameters as $\Delta_{l,a} = \Delta_{l,b} = \Delta_{l,Ta} = \Delta_{l,Tb} = 1, l = 1, \ldots, L$, the P_F^{local}s and P_M^{local}s for all the sensors are upper bounded by $2\tilde{\alpha}$ and $2\tilde{\beta}$, respectively. The underlying dependence or uncertainty resulting from the signals at the local sensors is decreased, and the signals received at the FC are near-independent. Compared to

the product-based SPRTs, our proposed copula-based SPRTs significantly improve the average expected stopping time, except for the distributed product-based SPRT of Case 1. This is also due to the near-independence (after local processing) among the received observations at the FC. Also, compared to the centralized cases, the distributed case requires more time to make a global decision. It seems that the expected average detection time of the distributed copula-based SPRT is much larger than that of the centralized copula-based SPRT. This is because we evaluate the performance at extremely low SNRs.

Case 2	P_F	P_M	$\mathbb{E}[T]$
Centralized product based SPRT	0.0067	0.0524	1.490
Centralized copula based SPRT	0.0054	0.0090	1.302
Distributed product based SPRT	0.0218	0.1184	14.843
Distributed copula based SPRT	0.0051	0.0100	4.020

Table 5.7: Unknown copula: Estimated P_F, P_M and $\mathbb{E}[T]$ with $\alpha = \beta = 0.01$ and SNR $= 0$ dB for Case 2.

To justify the efficiency of the distributed copula-based SPRT, in Table 5.7, we present the average P_F, P_M and $\mathbb{E}[T]$ for Case 2 at SNR $= 0$ dB. As we can see, the $\mathbb{E}[T]$ for the proposed distributed copula-based SPRT is quite comparable to the centralized copula-based SPRT and much faster than the distributed product-based SPRT.

5.6 Summary

In this chapter, we proposed a copula-based sequential scheme for the problem of distributed detection with imperfect communication channels from the sensors to the fusion center, where the sensor observations are assumed to be spatially dependent. This dependence may result from the dependent sensor signals, dependent channel noises or both. We used the regular

vine copula model to represent the underlying dependence. We first proposed a centralized copula-based SPRT, and showed its asymptotic optimality and time efficiency. We then proposed a distributed copula-based sequential scheme, where the memory-less truncated SPRTs were performed at the local sensors and the copula-based SPRT was conducted at the FC. We have shown that by suitably designing the local thresholds and the truncation window, the local probability of false alarm and the local probability of miss detection of the proposed memory-less truncated local sequential tests are upper bounded by the pre-specified error probabilities. Moreover, we have shown the asymptotic optimality and time efficiency of the distributed copula-based SPRT. Via simulations, we have shown that our proposed copula-based SPRTs can efficiently capture the unknown dependence, and outperform the product-based SPRTs which ignore the underlying dependence.

CHAPTER 6
DISTRIBUTED ESTIMATION IN LARGE SCALE WIRELESS SENSOR NETWORKS VIA A TWO-STEP CLUSTER-BASED APPROACH

6.1 Motivation

In large scale sensor networks, sensors observations are often assumed to be independent for analytical tractability. However, observations are often dependent in practical scenarios. Distributed estimation problems with independent sensor observations have been studied extensively (see e.g. [16, 20, 30, 31, 55, 56, 67, 68, 87]). However, the distributed estimation problem with dependent sensor observations has not received much attention. The underlying dependence can be both good and bad [44, 46, 115]. On the one hand, dependent sensors provide different viewpoints and aspects regarding the target parameter to be estimated. However, on the other hand, they may collect redundant observations. Therefore, spatial dependence needs

to be exploited properly to enhance the overall estimation efficiency. In [44, 46], the concepts of diversity gain and redundancy loss were introduced to characterize the influence of spatial dependence among sensor observations on estimation performance.

In this chapter, to achieve greater sensor transmission and estimation efficiencies, we propose a two-step cluster-based collaborative distributed estimation scheme. In the first step, sensors form dependence driven clusters such that the observations of sensors in the same cluster are dependent while the observations of sensors from different clusters are independent, and perform copula-based maximum a posteriori probability (MAP) estimation via intra-cluster collaboration. In the second step, the estimates generated in the first step are shared via inter-cluster collaboration to reach an average consensus. A merge based K-medoid dependence driven clustering algorithm is proposed. We further propose a cluster-based sensor selection scheme using mutual information prior to estimation. The aim is to select sensors with maximum relevance and minimum redundancy regarding the parameter of interest under certain pre-specified energy constraints.

6.2 Problem Formulation

Consider a phenomenon being observed by L sensors. Each sensor's observation is $z_l = \theta + w_l, \forall l = 1, \ldots, L$, where θ is the random parameter to be estimated corresponding to the phenomenon of interest and w_l is the observation noise which is spatially and temporally independent of θ. We assume that the prior distribution of θ is given as $f(\theta)$. Also, we assume that the observation noise can be dependent across some sensors. Moreover, we further assume that the sensor observations are continuous random variables that are conditionally independent and identically distributed (i.i.d.) over time. Let $f_l(\cdot|\theta)$ be the PDF of the observations at the lth sensor conditioned on θ. Note that the marginal conditional sensor PDFs can be distinct from each other. Throughout the chapter, we assume that given θ, the marginal distribution $f_l(\cdot|\theta), l = 1, \ldots, L$ is known.

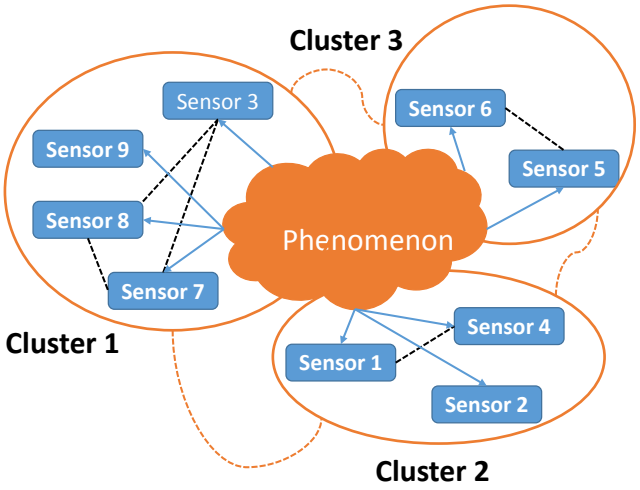

Fig. 6.1: Two-step cluster-based collaborative distributed estimation system, where the orange dash lines represent the inter-cluster communication links and the black dash lines denote the intra-cluster communication links.

In a non-collaborative setting, each sensor senses the phenomenon of interest and estimates the random parameter θ solely based on its own observations. In this work, we consider a two-step cluster-based collaborative scheme shown in Fig. 6.1, where in the first step, sensors form dependence driven clusters and extract information relevant for estimation by collaborating with other sensors in the same cluster. In the second step, local information obtained by each cluster in the first step is shared among clusters to yield a global estimate. The participating sensors are required to adhere to the following rules:

1. Sensors first form clusters, where each sensor is allowed to join only one cluster. The sensors that are most “similar", i.e., most statistically dependent, tend to stay in the same cluster.

2. Once the clusters are formed, a sensor can request observations from all the other sensors that are in the same cluster to perform estimation; it is also required to transmit its observations to the other collaborating sensors in the cluster based on their request.

3. A cluster can request the estimate of the parameter or observations from all the other clusters; it is also required to transmit its estimate of the parameter or observations to the

other collaborating clusters based on their request.

We denote the set of all the sensors in the network as $\mathcal{S}$, where the corresponding sensor observation set is $\mathbf{z}_{\mathcal{S}} = [\mathbf{z}_1, \ldots, \mathbf{z}_{|\mathcal{S}|}] \in \mathbb{R}^{N \times |\mathcal{S}|}$, $|\mathcal{S}| = L$, where $|\cdot|$ denotes the cardinality of a set and N is the number of observations for each sensor. Suppose there are K independent non-overlapping sensor clusters and denote the kth cluster by $\mathcal{G}_k, k \in [K]$, where for ease of notation, $[K]$ denotes $\{1, 2, \ldots, K\}$. Thus, $\mathcal{S} = \mathcal{G}_1 \cup \cdots \cup \mathcal{G}_K$.

In the estimation problem, Fisher Information (FI) is often used to characterize the amount of information that data carry about the parameter. It is given as

$$FI(\theta) = -\mathbb{E}_{\mathbf{x}} \left[\frac{\partial^2 \log f_{\mathbf{x}}(\mathbf{x}; \theta)}{\partial \theta^2} \right], \tag{6.1}$$

where $f_{\mathbf{x}}$ represents the joint PDF of the data sequence vector $\mathbf{x}$. For the entire sensor set $\mathcal{S}$, the FI it can achieve is given as

$$FI(\mathcal{S}) = -\mathbb{E}_{\mathbf{z}_{\mathcal{S}}} \left[\frac{\partial^2 \log f_{\mathbf{z}_{\mathcal{S}}}(\mathbf{z}_{\mathcal{S}}; \theta)}{\partial \theta^2} \right], \tag{6.2}$$

where $f_{\mathbf{z}_{\mathcal{S}}}$ is the joint distribution of $\mathbf{z}_{\mathcal{S}}$.

Proposition 6.1. *$FI(\mathcal{S})$ can be decomposed into cluster-based Fisher Information and prior Fisher Information.*

Proof:

$$
\begin{aligned}
FI(\mathcal{S}) &= -\mathbb{E}\left[\frac{\partial^2 \log\left(\prod_{l=1}^{L} f_l(\mathbf{z}_l|\theta) \times c_{\mathcal{S}}(\mathbf{F}(\mathbf{z}|\theta);\boldsymbol{\phi}) \times f(\theta)\right)}{\partial\theta^2}\right], \\
&= \sum_{l\in\mathcal{S}} FI_l + FI_p - \mathbb{E}\left[\frac{\partial^2 \log c_{\mathcal{S}}(\mathbf{F}(\mathbf{z}|\theta);\boldsymbol{\phi})}{\partial\theta^2}\right], \\
&\overset{(a)}{=} \sum_{l\in\mathcal{S}} FI_l + FI_p - \mathbb{E}\left[\frac{\partial^2 \log \prod_{k=1}^{K} c_{G_k}(\mathbf{F}(\mathbf{z}_{\mathcal{G}_k}|\theta);\boldsymbol{\phi}_k)}{\partial\theta^2}\right], \\
&= \sum_{l\in\mathcal{S}} FI_l + \sum_{k=1}^{K} FI_c(\mathcal{G}_k) + FI_p, \\
&= \sum_{k=1}^{K}\left(\sum_{l\in\mathcal{G}_k} FI_l + FI_c(\mathcal{G}_k)\right) + FI_p,
\end{aligned}
$$

where FI_p is the Fisher information with respect to the prior distribution on θ, (a) is obtained by using the assumption that sensor clusters are independent of each other. Also, we define $FI_c(\mathcal{G}_k)$ as $-\mathbb{E}\left[\frac{\partial^2 \log c_{G_k}(\mathbf{F}(\mathbf{z}_{\mathcal{G}_k}|\theta);\boldsymbol{\phi}_k)}{\partial\theta^2}\right]$.

Therefore, $FI(\mathcal{S})$ can be decomposed into cluster-based Fisher Information and prior Fisher Information. ■

Remark 6.1. *Based on Proposition 6.1, if the sensors in the network can be perfectly clustered into independent non-overlapping clusters, we can process each cluster independently and then combine each cluster's information to obtain the global estimate.*

In the first step, an intuitive solution would be that each cluster learns its dependence structure, and shares the estimated conditional joint PDFs with all the other clusters in the second step. Then, the estimation problem becomes

$$
\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{N}\sum_{k=1}^{K} \left(\log f(\mathbf{z}_{\mathcal{G}_k,i}|\theta)\right) + \log f(\theta), \tag{6.3}
$$

where $\mathbf{z}_{\mathcal{G}_k,i}$ is the observation set for cluster $\mathcal{G}_k$ at time instant i and $f(\mathbf{z}_{\mathcal{G}_k}|\theta)$ is the conditional

joint PDF of the sensor observations (which is not known a *priori*) in cluster $\mathcal{G}_k, k \in [K]$. $f(\theta)$ is the prior distribution of θ.

Remark 6.2. *The estimation methodology given in Equation* (6.3) *is referred to as cluster-based MAP scheme. Note that the conditional joint PDF* $f(\mathbf{z}_{\mathcal{G}_k}|\theta)$ *in Equation* (6.3) *can be estimated using copula based methods that take dependent observations into consideration (see Equation* (1.2)*). The cluster-based MAP scheme using copula incorporated approach is optimal.*

However, transmitting the estimated conditional joint PDFs and the raw observations among clusters can be expensive. Therefore, we propose to share estimates obtained by each cluster until a consensus is achieved.

In the following section, we present the details of our two-step cluster-based distributed estimation scheme, including the clustering of the sensors, the intra-cluster collaborative estimation approach using copula based methods and the inter-cluster collaboration strategy.

6.3 Two-Step Dependence Driven Collaborative Distributed Estimation

In this section, we present our two-step cluster-based collaborative distributed estimation scheme. In the first step, sensors form clusters based on their similarity/dissimilarity with the other sensors. We propose a merge based K-medoid dependence driven clustering algorithm. After the clusters are formed, each sensor then estimates θ using copula based MAP via intra-cluster collaboration. In the second step, the estimated θs are shared among clusters to yield a consensus. Here, we assume that the sensors and the sensor clusters communicate via error-free, orthogonal channels. Before we proceed, we first make some assumptions and define the dissimilarity measures.

6.3.1 Assumptions and Dissimilarity Measure Definitions

We define the inter-cluster dissimilarity between $\mathcal{G}_k$ and $\mathcal{G}_{k'}$ as well as the intra-cluster dissimilarity of $\mathcal{G}_k$, respectively as

$$d(\mathcal{G}_k, \mathcal{G}_{k'}) = \inf_{s_i \in \mathcal{G}_k, s_j \in \mathcal{G}_{k'}} d(s_i, s_j),$$

$$d(\mathcal{G}_k) = \sup_{s_i, s_j \in \mathcal{G}_k} d(s_i, s_j),$$

where $d(\cdot, \cdot)$ is a dissimilarity metric between two variables/data sequences, e.g., the rank based dissimilarity measure defined later in Equation (6.4). Here, $d(\mathcal{G}_k, \mathcal{G}_{k'})$ represents the dissimilarity between cluster $\mathcal{G}_k$ and $\mathcal{G}_{k'}$. We further define

$$d_H = \min_{k,k'=1,\ldots,K, k \neq k'} d(\mathcal{G}_k, \mathcal{G}_{k'}),$$

$$d_L = \max_{k=1,\ldots,K} d(\mathcal{G}_k).$$

We make the following assumptions:

A_1 $d_L < d_H$,

A_2 $P\left(d(\mathbf{z}_i^k, \mathbf{z}_j^{k'}) \leq d_0\right) < \epsilon_1, d_0 \in (d_L, d_H)$,

A_3 $P\left(d(\mathbf{z}_i^k, \mathbf{z}_j^k) > d_0\right) < \epsilon_2, d_0 \in (d_L, d_H)$,

A_4 $P\left(d(\mathbf{z}_i^k, \mathbf{z}_j^k) \geq d(\mathbf{z}_i^k, \mathbf{z}_{j'}^{k'})\right) < \epsilon_3$,

where $\mathbf{z}_i^k, \mathbf{z}_j^k \in \mathbf{z}_{\mathcal{G}_k}$ and $\mathbf{z}_j^{k'}, \mathbf{z}_{j'}^{k'} \in \mathbf{z}_{\mathcal{G}_{k'}}$. $\epsilon_i > 0, i = 1, 2, 3$ are small constants and N is the number of observations.

Assumption A_2 implies that the probability that the dissimilarity between sensor observations obtained from two different clusters is smaller than d_H is small. Also, assumption A_3 guarantees that the probability that the dissimilarity between sensor observations obtained

from the same cluster is greater than d_L is small. Assumption A_4 states that given two sensor observation sequences generated from the same cluster and a third observation sequence generated from another cluster, the probability that the first sequence is closer to the third sequence is small. Due to the use of measure noisy data, the assumptions A_1 to A_4 imply that sensors that are from the same cluster are dependent, while the ones that are from different clusters are nearly independent.

The dissimilarity between two sensors can be characterized using different dependence measures, such as the Pearson's correlation coefficient , a rank based correlation measure (Spearman's ρ and Kendall's τ) and the copula based measure. Note that the Pearson's correlation coefficient that characterizes linear relationship is inadequate to capture nonlinear dependence among the involved sensors. Also, the copula based measure is not a symmetric dependence measure. In the following, we propose a dissimilarity metric based on rank based correlation.

Let $\kappa \in [-1, 1]$ be a rank based measure (Spearman's ρ or Kendall's τ). We introduce a dissimilarity function $d(\mathbf{x}, \mathbf{y})$ between the random variables X and Y, where $\mathbf{x} = [x_1, \ldots, x_N]$ and $\mathbf{y} = [y_1, \ldots, y_N]$ are the i.i.d. data sequences corresponding to the variables X and Y, respectively, given as

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{1 - \kappa(\mathbf{x}, \mathbf{y})^2}, \tag{6.4}$$

where N is the number of samples for variables X and Y and $\kappa(\mathbf{x}, \mathbf{y})$ is Spearman's ρ or Kendall's τ between sequences $\mathbf{x}$ and $\mathbf{y}$. Note that if $\kappa(\mathbf{x}, \mathbf{y}) = 1$ or $\kappa(\mathbf{x}, \mathbf{y}) = -1$, we have $d(\mathbf{x}, \mathbf{y}) = 0$.

6.3.2 Dependence Driven Clustering Process

We propose a dependence driven clustering scheme. Let $d(\mathbf{z}_i, \mathbf{z}_j)$ denote the dissimilarity between the ith and jth sensors, where $i, j \in [L]$. Therefore, $d(\mathbf{z}_i, \mathbf{z}_j)$ is small when sensor i and sensor j are strongly dependent and is large when sensor i and sensor j are weakly

dependent.

The goal of the clustering process is to cluster the sensors in the network based on the underlying dependence among sensors. The number of clusters K is unknown. Therefore, we need to estimate it. Here, we apply a merge based K-medoid clustering scheme [6,44,46,113] to perform the clustering and find $\hat{K}$. The merging criterion is that if the dissimilarity/distance of any two clusters is greater than $d_{th} \in (d_L, d_H)$, these two clusters should be separated; otherwise, they merge together.

In the following, we present the initialization of the cluster centers and clusters. Before, we initialize the clusters, the centers need to be initialized first. We denote the cluster centers as $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{\hat{K}}$ and the cluster center set as $\boldsymbol{\mu}$. We first arbitrarily choose $\mathbf{z}_i, i \in [L]$ as $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu} = \{\boldsymbol{\mu}_1\}$. Then, for $\max\limits_{\mathbf{z}_i \in \mathbf{z}_{\mathcal{S}} \setminus \boldsymbol{\mu}} \left(\min\limits_{\boldsymbol{\mu}_k \in \boldsymbol{\mu}} d\left(\mathbf{z}_i, \boldsymbol{\mu}_k\right) \right) > d_{th}$, we do

$$\begin{aligned} \tilde{\boldsymbol{\mu}} &= \arg\max_{\mathbf{z}_i \in \mathbf{z}_{\mathcal{S}} \setminus \boldsymbol{\mu}} \left(\min_{\boldsymbol{\mu}_k \in \boldsymbol{\mu}} d\left(\mathbf{z}_i, \boldsymbol{\mu}_k\right) \right), \\ \boldsymbol{\mu} &= \boldsymbol{\mu} \cup \tilde{\boldsymbol{\mu}}. \end{aligned} \tag{6.5}$$

After we obtain the cluster centers, the clusters, which are originally defined as empty sets, are initialized as: for $i = 1, 2, \ldots, L$

$$\begin{aligned} \boldsymbol{\mu}_j &= \arg\min_{\mu_j \in \boldsymbol{\mu}} d\left(\mathbf{z}_i, \boldsymbol{\mu}_j\right), \\ \mathcal{G}_j &\leftarrow \mathcal{G}_j \cup \{\mathbf{z}_i\}. \end{aligned} \tag{6.6}$$

The proposed dependence driven clustering scheme is shown in Algorithm 6.1.

6.3.3 Copula Based MAP

After the clusters are formed, each sensor performs estimation by collaborating with the sensors in the same cluster. We assume a fully connected network for intra-cluster collaboration.

Algorithm 6.1 Dependence Driven Clustering.

Input: Sensor observations $\{\mathbf{z}_1, \ldots, \mathbf{z}_L\}$ and threshold d_{th}.
Output: Clusters $\{\mathcal{G}_k\}_{k=1}^{\hat{K}}$.

1. Initialize clusters $\{\mathcal{G}_k\}_{k=1}^{\hat{K}}$
2. **while** not converge **do**
3. **Center update:**
4. **for** $k = 1$ to $\hat{K}$ **do**

$$\boldsymbol{\mu}_k \leftarrow \arg\min_{\mathbf{z}_l \in \mathcal{G}_k} \sum_{\mathbf{z}_{l'} \in \mathcal{G}_k} d\left(\mathbf{z}_l, \mathbf{z}_{l'}\right)$$

5. **end for**
6. **Merge step:**
7. **for** $k_1, k_2 \in [1, 2, \ldots, \hat{K}]$ and $k_1 \neq k_2$ **do**
8. **if** $d(\boldsymbol{\mu}_{k_1}, \boldsymbol{\mu}_{k_2}) \leq d_{th}$ **then**
9. **if** $\sum_{\mathbf{z}_l \in \mathcal{G}_{k_1}} d(\boldsymbol{\mu}_{k_2}, \mathbf{z}_l) < \sum_{\mathbf{z}_l \in \mathcal{G}_{k_2}} d(\boldsymbol{\mu}_{k_1}, \mathbf{z}_l)$
10. **then** $\mathcal{G}_{k_2} \leftarrow \mathcal{G}_{k_1} \cup \{\mathcal{G}_{k_2}\}$ and delete $\boldsymbol{\mu}_{k_1}$ and $\mathcal{G}_{k_1}$
11. **else** $\mathcal{G}_{k_1} \leftarrow \mathcal{G}_{k_1} \cup \{\mathcal{G}_{k_2}\}$ and delete $\boldsymbol{\mu}_{k_2}$ and $\mathcal{G}_{k_2}$
12. **end if**
13. $\hat{K} \leftarrow \hat{K} - 1$
14. **end if**
15. **end for**
16. **Cluster update:**
17. **for** $l = 1$ to L **do**
18. **if** $\mathbf{z}_l \in \mathcal{G}_{k'}$ and $d\left(\mathbf{z}_l, \boldsymbol{\mu}_k\right) < d\left(\mathbf{z}_l, \boldsymbol{\mu}_{k'}\right)$ then

$$\mathcal{G}_k \leftarrow \mathcal{G}_k \cup \{\mathbf{z}_l\} \text{ and } \mathcal{G}_{k'} \leftarrow \mathcal{G}_{k'} \setminus \{\mathbf{z}_l\}$$

19. **end if**
20. **end for**
21. **end while**
22. **Return** $\{\mathcal{G}_k\}_{k=1}^{\hat{K}}$

In each cluster, each sensor estimates θ using MAP based on its own observations and observations from all the other collaborating sensors in the same cluster. Note that for a fully connected network, all the sensors in the cluster have the same set of observations available to them. We denote the corresponding sensor observations for cluster $\mathcal{G}_k$ as $\mathbf{z}_{\mathcal{G}_k} = \{\mathbf{z}_1^k, \mathbf{z}_2^k, \ldots, \mathbf{z}_{|\mathcal{G}_k|}^k\}$. Therefore, the estimate $\hat{\theta}_k$ at each sensor for the kth cluster is given by

$$\hat{\theta}_k = \arg\max_{\theta} \sum_{i=1}^{N} \log\left(f(z_{1i}^k, z_{2i}^k, \ldots, z_{|\mathcal{G}_k|i}^k|\theta) \times f(\theta)\right), \tag{6.7}$$

where N is the number of observations and $f(z_{1i}^k, z_{2i}^k, \ldots, z_{|\mathcal{G}_k|i}^k; \theta)$ is the joint PDF of all the sensors in cluster $\mathcal{G}_k, k \in [K]$ at time instant $i, i \in [N]$.

We use the copula based approach to characterize the underlying dependence in each cluster and according to Equation (1.2), $\hat{\theta}_k$ can be obtained by

$$\begin{aligned}\hat{\theta}_k = \arg\max_{\theta} \sum_{i=1}^{N} \sum_{l=1}^{|\mathcal{G}_k|} \log f_l(z_{li}^k|\theta) + \sum_{i=1}^{N} \log c_k(\mathbf{F}(\mathbf{z}_i^k|\theta); \boldsymbol{\phi}_k) \\ + \log f(\theta),\end{aligned} \tag{6.8}$$

where $\mathbf{F}(\mathbf{z}_i^k|\theta) = [F(z_{1i}^k|\theta), F(z_{2i}^k|\theta), \ldots, F(z_{|\mathcal{G}_k|i}^k|\theta)]$ is the set of marginal CDFs, and $c_k(\cdot; \boldsymbol{\phi}_k)$ is the multivariate copula density function and $\boldsymbol{\phi}_k$ is the corresponding parameter set for cluster $k, k \in [K]$.

Typically, the multivariate dependence $c_k(\cdot; \boldsymbol{\phi}_k)$ in Equation (6.8) is unknown a *priori* and needs to be estimated. Since the learning of the copula models is similar for all the clusters, in the following, we omit the cluster index k for simplification of notation.

To estimate the multivariate copula $c(\cdot; \boldsymbol{\phi})$, we first define a library of copula models, $\mathcal{C} = \{c_j : j = 1, \ldots, M\}$ [75]. The optimal copula model is then determined by the Akaike Information Criterion (AIC) [3] in Equation (6.12), namely, the best copula is the copula model with minimum AIC value. Before evaluating the AIC values for each copula model, we need

to estimate the marginal CDFs and the associated copula parameter(s) $\boldsymbol{\phi}_j, j = 1, \ldots, M$. The marginal CDFs can be estimated using EPIT [45]:

$$\hat{F}_l(x) = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{I}(z_{li} < x), \tag{6.9}$$

where $\boldsymbol{I}$ is the indicator function and N is the number of observations for estimation. The copula parameter(s) $\boldsymbol{\phi}_j$ can then be estimated using MLE, which is given by

$$\widehat{\boldsymbol{\phi}}_j = \arg\max_{\boldsymbol{\phi}_j} \sum_{i=1}^{N} \log c_j(\hat{F}_{1i}, \ldots, \hat{F}_{|\mathbf{s}_{m_k}|i} | \boldsymbol{\phi}_j). \tag{6.10}$$

With the estimated parameter(s), the best copula c^* is given as

$$c^* = \arg\min_{c_j \in \mathcal{C}} \text{AIC}_j. \tag{6.11}$$

The AIC value is given as

$$\text{AIC}_i = -\sum_{n=1}^{N} \log c_i(\hat{F}_{1n}, \ldots, \hat{F}_{|\mathbf{s}_{m_k}|n} | \widehat{\boldsymbol{\phi}}_i) + q(c_i), \tag{6.12}$$

where $q(c_i)$ is the number of parameters in the ith copula model.

6.3.4 Cluster Based Consensus Scheme

After all the clusters obtain their initial estimates, these estimates are shared via linear inter-cluster collaboration to reach a consensus. We employ the average consensus algorithm [80]. Assume that the collaboration among clusters is represented by a fixed topology matrix $\mathbf{A}$ with binary entries, namely, $A_{ij} \in \{0, 1\}, i, j \in [K]$. Here, $A_{ij} = 1$ means that there is a communication link from the ith cluster to the jth cluster; otherwise, $A_{ij} = 0$. At iteration

$n+1$, each cluster $\mathcal{G}_i, i \in [K]$ updates its estimate $\hat{\theta}_{\mathcal{G}_i}(n+1)$ as follows [80]:

$$\hat{\theta}_{\mathcal{G}_i}(n+1) = \hat{\theta}_{\mathcal{G}_i}(n) - \beta \sum_{j \in N_{\mathcal{G}_i}} A_{ij} \left(\hat{\theta}_{\mathcal{G}_j}(n) - \hat{\theta}_{\mathcal{G}_i}(n) \right), \tag{6.13}$$

where $0 < \beta < 1/\Delta$, Δ is the maximum degree of the network and $N_{\mathcal{G}_i}$ is the neighborhood cluster set of $\mathcal{G}_i$.

It has been shown in [80, Theorem 2] that if the graph is strongly connected and balanced, $\hat{\theta} = \frac{\sum_i \hat{\theta}_{\mathcal{G}_i}(0)}{K}$ asymptotically.

Theorem 6.1. *The standard deviation of the parameter estimate obtained by the average consensus scheme is upper bounded by the average standard deviation of all the clusters' estimates, i.e.,*

$$\sqrt{var(\hat{\theta})} \leq \frac{1}{K} \sum_{k=1}^{K} \sqrt{var(\hat{\theta}_k)} \tag{6.14}$$

where $k, k \in [K]$ denotes the cluster index and $var(\cdot)$ represent the variance of a random variable.

Proof: See Appendix F. ■

Remark 6.3. *The average consensus based inter-cluster collaboration helps in mitigating the effect of estimation bias resulting from the individual cluster estimates.*

Since the sensor network is large, the number of sensors in each cluster is also potentially large. As mentioned in Section 6.1, some sensors may provide redundant information. Allowing all the sensors in the cluster to exchange their information may result in a large transmission cost. Therefore, selecting sensors with maximum information and minimum redundancy is crucial. In the following section, we propose a mutual information based sensor selection scheme, and only the selected sensors need to exchange their information within a cluster.

Remark 6.4. *In practice, to extend the network lifetime, one may design sleep scheduling schemes for sensors which provide redundant data [25, 26]. Also, to balance battery usage*

for inter-cluster communication, one may rotate sensors that are responsible for inter-cluster collaboration in a small region near the edge of the cluster.

6.4 Sensor Selection Based Two-Step Dependence Driven Collaborative Distributed Estimation

In this section, we present the details of the sensor selection scheme for our two-step collaborative estimation scheme. For each cluster $\mathcal{G}_k, k \in [K]$, prior to estimation via intra-cluster collaboration, a mutual information based methodology is employed to select sensors with maximum information and minimum redundancy.

Before we proceed, we recall that the mutual information of two random variables x and y, denoted by $I(x;y)$, is given as

$$I(x;y) = \int f(x,y) \log \left(\frac{f(x,y)}{f(x)f(y)} \right) dxdy, \tag{6.15}$$

where $f(x,y)$ is the joint PDF of variables x and y. $f(x)$ and $f(y)$ are the marginal PDFs.

The optimal sensor selection strategy is often based on maximal relevance and minimal redundancy with respect to the target parameter θ on the entire sensor set [83], and this strategy is referred as *maximal-relevancy-minimal-redundancy* (mRMR) in [83]. Suppose that we aim to select m sensors from the set of all the sensors in the network $\mathcal{S}$ with the corresponding observation set $\mathbf{z}_{\mathcal{S}} = \{\mathbf{z}_1, \ldots, \mathbf{z}_{|\mathcal{S}|}\}$. In terms of mutual information, the mRMR solution is obtained by solving the following problem

$$\max_{\mathbf{s}_m \in \mathcal{S}} \left[\frac{1}{|\mathbf{s}_m|} \sum_{\mathbf{z}_i \in \mathbf{z}_{\mathbf{s}_m}} I(\mathbf{z}_i;\theta) - \frac{1}{|\mathbf{s}_m|^2} \sum_{\mathbf{z}_i, \mathbf{z}_j \in \mathbf{z}_{\mathbf{s}_m}} I(\mathbf{z}_i;\mathbf{z}_j) \right], \tag{6.16}$$

where $\mathbf{s}_m$ is the set of the selected sensors with cardinality $|\mathbf{s}_m| = m$ and $\mathbf{z}_{\mathbf{s}_m} = \{\mathbf{z}_1, \ldots, \mathbf{z}_{|\mathbf{s}_m|}\}$

is the sensor observation set of $\mathbf{s}_m$, where $\mathbf{z}_{\mathbf{s}_m} \in \mathbf{z}_{\mathcal{S}}$.

Note that the computational complexity of the mRMR problem is $O(|\mathcal{S}|^m)$. A more efficient first-order incremental search method was proposed to find the near-optimal solutions of problem in Equation (6.16) in [83]. It is given as:

$$\max_{\mathbf{z}_j \in \mathbf{z}_{\mathcal{S}} \setminus \mathbf{z}_{\mathbf{s}_{m-1}}} \left[I(\mathbf{z}_j; \theta) - \frac{1}{m-1} \sum_{\mathbf{z}_i \in \mathbf{z}_{\mathbf{s}_{m-1}}} I(\mathbf{z}_j; \mathbf{z}_i) \right], \tag{6.17}$$

where $\mathbf{s}_{m-1}$ is the selected sensor set with $m-1$ sensors, and $\mathbf{z}_{\mathcal{S}} \setminus \mathbf{z}_{\mathbf{s}_{m-1}}$ denotes that we exclude the sensor observations from the sensors in set $\mathbf{s}_{m-1}$ from $\mathbf{z}_{\mathcal{S}}$.

The computational complexity of the incremental search method in Equation (6.17) is $O(m * |\mathcal{S}|)$. To further reduce the computational complexity, in the following, we propose a cluster-based incremental search methodology, where the sensor selection is performed cluster-by-cluster independently.

Note that $\mathcal{S} = \{\mathcal{G}_1 \cup \mathcal{G}_2 \ldots \cup \mathcal{G}_K\}$, where $\mathcal{G}_i \cap \mathcal{G}_j = \emptyset, i, j \in [K]$. Instead of searching over the entire sensor set, we select m_k sensors from cluster $\mathcal{G}_k, k \in [K]$. Note that $\sum_{k=1}^{K} m_k = m$.

For each cluster $\mathcal{G}_k$, suppose that we already have $\mathbf{s}_{m_k-1}$, the sensor set with $m_k - 1$ sensors. The incremental selection scheme solves the following problem:

$$\max_{\mathbf{z}_j \in \mathbf{z}_{\mathcal{G}_k} \setminus \mathbf{z}_{\mathbf{s}_{m_k-1}}} \left[I(\mathbf{z}_j; \theta) - \frac{1}{m_k-1} \sum_{\mathbf{z}_i \in \mathbf{z}_{\mathbf{s}_{m_k-1}}} I(\mathbf{z}_j; \mathbf{z}_i) \right], \tag{6.18}$$

where $\mathbf{z}_{\mathcal{G}_k} \setminus \mathbf{z}_{\mathbf{s}_{m_k-1}}$ denotes that we exclude the sensor observations in set $\mathbf{s}_{m_k-1}$ from $\mathbf{z}_{\mathcal{G}_k}$.

Remark 6.5. *The selection scheme given in Equation* (6.17) *is referred to as the global sensor selection scheme. Also, the selection scheme given in Equation* (6.18) *is referred to as the cluster-based sensor selection scheme.*

Theorem 6.2. *Using a suitably designed threshold d_{th} that makes inter-cluster sensors nearly independent, the cluster-based sensor selection method is equivalent to the global sensor se-*

lection method with probability at least $1-\epsilon$*, where* ϵ *is a small constant.*

Proof: See Appendix G. ■

A natural question is how to determine the optimal number of sensors m_k for cluster $\mathcal{G}_k, k \in [K]$. In an energy constrained network with battery-limited sensors, each sensor's energy is finite and a communication cost is incurred when it transmits observations to collaborating sensors. Therefore, the number of sensors that can be selected in each cluster is limited due to finite energy budgets. Let r be the average number of requests initiated by each sensor in the network per unit time interval. Then, for the selected sensors in cluster $\mathcal{G}_k$, the number of requests that have to be responded to within a unit time interval is $r(m_k-1)$. Moreover, we assume that the energy cost for a single transmission is E_t. The average energy consumption per unit time interval for each selected sensor in cluster $\mathcal{G}_k$ is $\mathbb{E}[s] = r(m_k-1)E_t, s \in \mathbf{s}_{m_k}$, which increases as the size of the selected sensor set $\mathbf{s}_{m_k}$ increases. Let the energy consumption of cluster $\mathcal{G}_k$ be the average energy consumption per sensor in $\mathbf{s}_{m_k}$, denoted by $\mathbb{E}[\mathbf{s}_{m_k}]$. Thus, in terms of energy efficiency, a smaller sensor set is preferred. In order to guarantee adequate sensors lifetimes, we enforce the energy consumption constraint as follows:

$$\mathbb{E}[\mathbf{s}_{m_k}] = r(m_k-1)E_t \leq \alpha_k, k \in [K], \tag{6.19}$$

where $\alpha_k > 0$ is the pre-specified constraint for cluster k. Therefore, the energy constrained selection scheme for each cluster $\mathcal{G}_k, k \in [K]$ is stated as

$$\begin{aligned} \max_{\mathbf{s}_{m_k} \in \mathcal{G}_k} \quad & \frac{1}{m_k} \sum_{\mathbf{z}_i \in \mathbf{z}_{\mathbf{s}_{m_k}}} I(\mathbf{z}_i;\theta) - \frac{1}{m_k^2} \sum_{\mathbf{z}_i, \mathbf{z}_j \in \mathbf{z}_{\mathbf{s}_{m_k}}} I(\mathbf{z}_i, \mathbf{z}_j), \\ \text{subject to} \quad & \mathbb{E}[\mathbf{s}_{m_k}] \leq \alpha_k. \end{aligned} \tag{6.20}$$

The problem in Equation (6.20) can be solved using the incremental search method in Equation (6.18) while satisfying the energy constraint.

6.5 Simulation Results

In this section, we demonstrate the efficacy of our proposed two-step cluster-based collaborative distributed estimation methodologies through numerical examples. We consider a wireless sensor network with $L = 13$ sensors deployed in a $[0, 1.5] \times [0, 1.5]$ square area of interest. Let (x_0, y_0) be the target location coordinates and θ be the intensity of the target signal to be estimated. We assume a Gaussian prior $\mathcal{N}(\theta_p, \sigma_p^2)$ on θ. Sensor $l, l = 1, \ldots, L$ is located at (x_l, y_l). The received measurements at the lth sensor are modeled as

$$z_{li} = \theta + w_{li}, i = 1, \ldots, N, \tag{6.21}$$

where w_{li} is the measurement noise which is assumed to be Gaussian distributed with mean 0 and variance σ_l^2 and N is the number of observations. Here, we assume that the variance of the measurement noise at each sensor is inversely scaled by the distance between the sensor and the signal source, i.e., $\sigma_l^2 = \frac{\sigma_0^2}{\sqrt{(x_l - x_0)^2 + (y_l - y_0)^2}}$. Note that σ_0^2 is introduced here for the ease of characterizing signal to noise ratio (SNR) at different sensors. We define our SNR as

$$\text{SNR} = \frac{\mathbb{E}[\theta^2]}{\sigma_0^2}. \tag{6.22}$$

We assume that the measurement noise w_{li} and θ are independent of each other. Moreover, we assume that the measurement noises are i.i.d. across time and can be spatially dependent at some sensors. Without loss of generality, we assume that we have three clusters and the underlying spatial dependence among sensors is generated cluster by cluster using multivariate Clayton copula functions. The pair-wise sensor dissimilarities are estimated based on Kendall's τ. We set $rE_t = 1$. Therefore, according to Equation (6.19), the maximum number of sensors that can be selected in cluster $k, k \in [K]$ is $m_k = \alpha_k$. Also, without loss of generality, we assume that $\alpha_1 = \alpha_2 = \ldots, \alpha_K$. Therefore, $m_1 = m_2 = \ldots, m_K$. The total number of sensors that are selected is $m = \sum_{k=1}^{K} m_k$.

We use average mean squared error (MSE) to characterize the estimation performance. For the clustering process, we use the average clustering accuracy to measure the clustering performance. The clustering accuracy is defined as $\frac{\text{Number of correctly clustered sensors}}{\text{Total number of sensors}}$. All the results are obtained using 500 Monte Carlo trials.

To exhibit the performance improvement by applying our proposed two-step cluster-based collaborative distributed estimation methodologies, we also evaluate the corresponding estimation performance under independence assumption that ignores dependence among sensor observations. Moreover, we compare our proposed estimation methodologies with the cluster-based MAP method given in Equation (6.3), where the copula-based approach as well as the product-based approach (under independence assumption) can be used to model the conditional joint PDFs. For clarity, we summarize the eight empirically studied cases as follows.

- Cluster-based consensus with sensor selection using copula based method as well as under independence assumption
- Cluster-based consensus without sensor selection using copula based method as well as under independence assumption
- Cluster-based MAP with sensor selection using copula based method as well as under independence assumption
- Cluster-based MAP without sensor selection using copula based method as well as under independence assumption

In Fig. 6.2, we present the average clustering accuracy as a function of the threshold d_{th} at SNR $= 2.0$ dB and $N = 70$. We can see that the choice of d_{th} has a significant impact on the performance of Algorithm 6.1. The optimal value of d_{th} depends on the given data, namely, d_L and d_H. Moreover, as we can see, a larger d_{th} results in a better clustering performance.

In Fig. 6.3, we present the the average clustering accuracy as a function of the number of observations N with $d_{th} = 0.83$ at SNR $= 2.0$ dB. As we can see, by choosing appropriate d_{th}

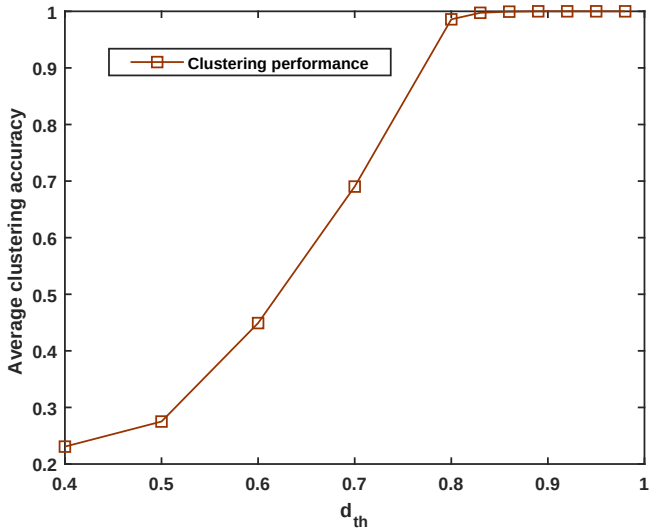

Fig. 6.2: Average clustering accuracy as a function of threshold d_{th}.

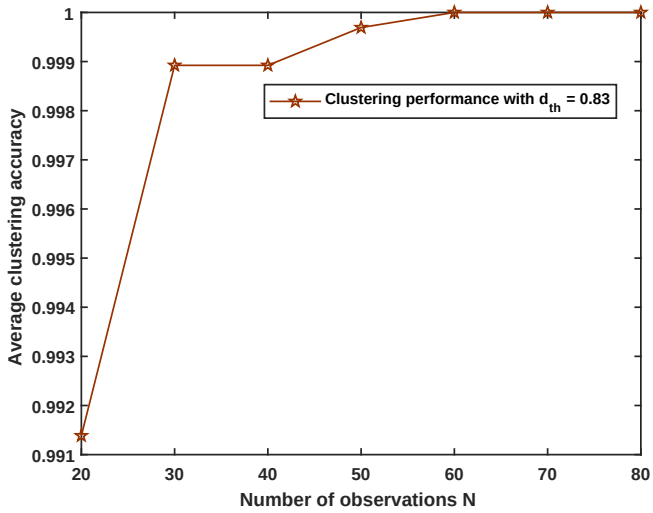

Fig. 6.3: Average clustering accuracy as a function of number of observations N with $d_{th} = 0.83$.

and N, we can achieve perfect clustering performance. In the following, our estimation results are obtained with $d_{th} = 0.83$ and $N = 70$ unless otherwise specified.

In Fig. 6.4 and Fig. 6.5, we present the average MSE as a function of SNR and the number of observations N, respectively, and compare the performance of schemes without sensor selection. The schemes that are evaluated are: Cluster-based consensus without sensor selection

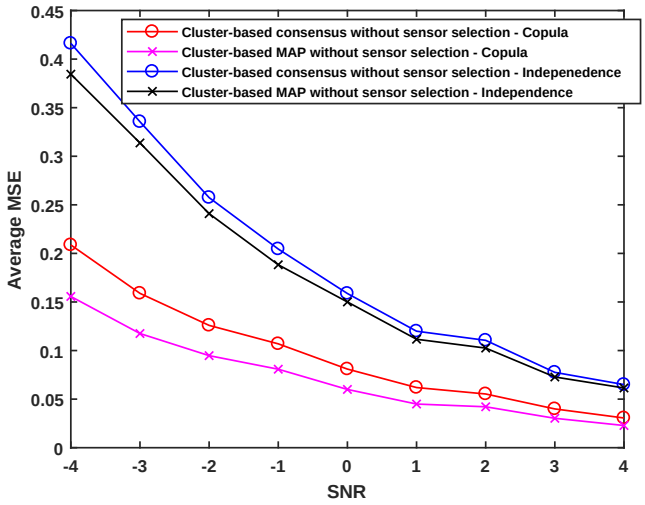

Fig. 6.4: Average MSE as a function of SNR without sensor selection.

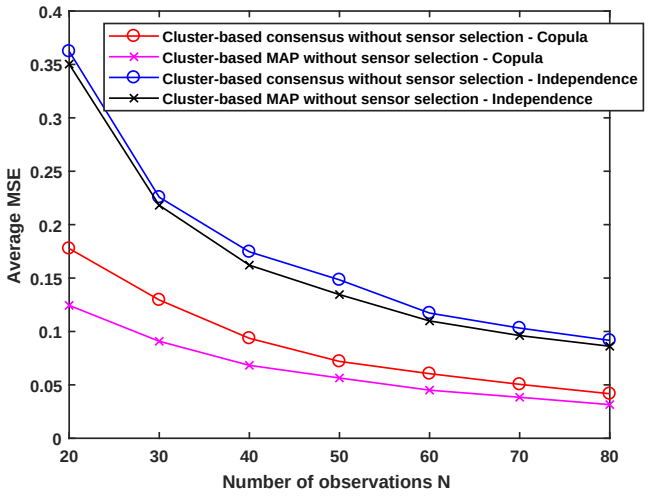

Fig. 6.5: Average MSE as a function of the number of observations N without sensor selection.

using copula based scheme, Cluster-based MAP without sensor selection using copula based scheme, Cluster-based consensus without sensor selection under independence assumption and Cluster-based MAP without sensor selection under independence assumption. We can see that as N as well as SNR increases, the average MSE decreases. Also, the schemes using copula based estimation methodologies perform significantly better than the schemes that assume independence among sensor observations. Moreover, as we can see, for the independent cases, the cluster-based consensus scheme performs pretty close to the cluster-based MAP scheme

while for the copula cases, the cluster-based consensus scheme performs close to the corresponding cluster-based MAP scheme at the SNR values greater than 0 dB in Fig. 6.4 and the number of observations larger than 50 in Fig. 6.5. Note that with extremely low SNR values or very small number of observations, the estimation performance difference between the copula incorporated cluster-based consensus scheme and the copula incorporated cluster-based MAP scheme is large. This is because for the cluster-based consensus scheme, the estimate obtained from each cluster is relatively poor for extremely low SNR values or with very small number of observations while for the cluster-based MAP scheme, it models the conditional joint PDF and captures more information.

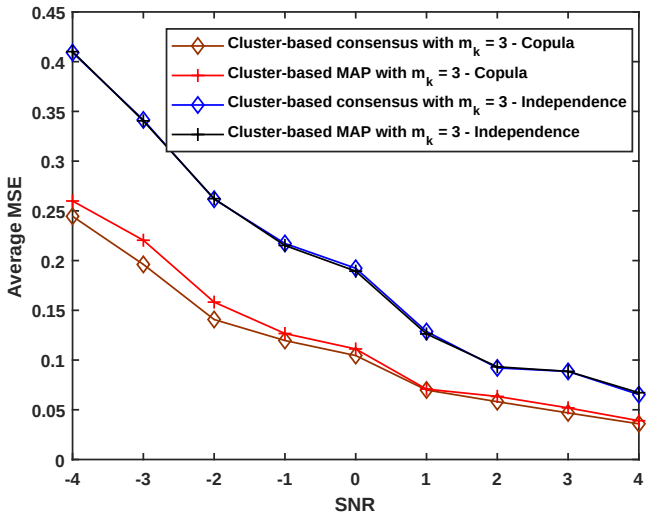

Fig. 6.6: Average MSE as a function of SNR with cluster-based sensor selection and $m_k = 3$.

In Fig. 6.6 and Fig. 6.7, we present the average MSE as a function of SNR and the number of observations N, respectively, by comparing schemes with cluster-based sensor selection. The schemes that are evaluated are: Cluster-based consensus with $m_k = 3$ using copula based scheme, Cluster-based MAP with $m_k = 3$ using copula based scheme, Cluster-based consensus with $m_k = 3$ under independence assumption and Cluster-based MAP with $m_k = 3$ under independence assumption. As we can see, the schemes using copula based estimation methodologies perform significantly better than the schemes assuming independence among sensor

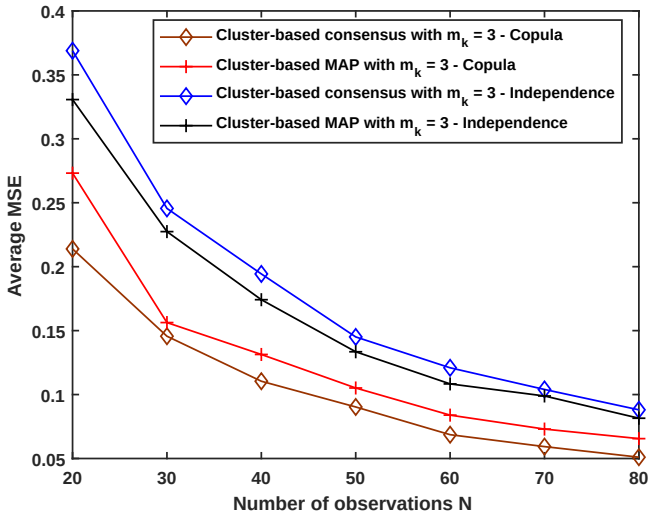

Fig. 6.7: Average MSE as a function of the number of observations N with cluster-based sensor selection and $m_k = 3$.

observations. Note that with sensor selection, the cluster-based consensus scheme using copula incorporated estimation methodology performs better than the corresponding cluster-based MAP scheme. This is because our proposed sensor selection scheme aims to select sensors with maximum relevance and minimum redundancy (namely, most independent sensors) regarding the parameter of interest. With most independent selected sensors, part of the dependence information for each cluster is lost. For the cluster-based MAP scheme, the product approach is used to combine the conditional joint PDFs corresponding to each cluster whereas for the cluster-based consensus scheme, consensus is used and the estimates obtained from each cluster are linearly combined where the linear dependence is imposed inherently resulting in better performance.

In Fig. 6.8 and Fig. 6.9, we present the average MSE as a function of SNR and the number of observations N, respectively, for copula incorporated schemes with cluster-based sensor selection and the copula incorporated schemes without sensor selection. The schemes that are evaluated are: Cluster-based consensus without sensor selection using copula based scheme, Cluster-based MAP without sensor selection using copula based scheme, Cluster-based con-

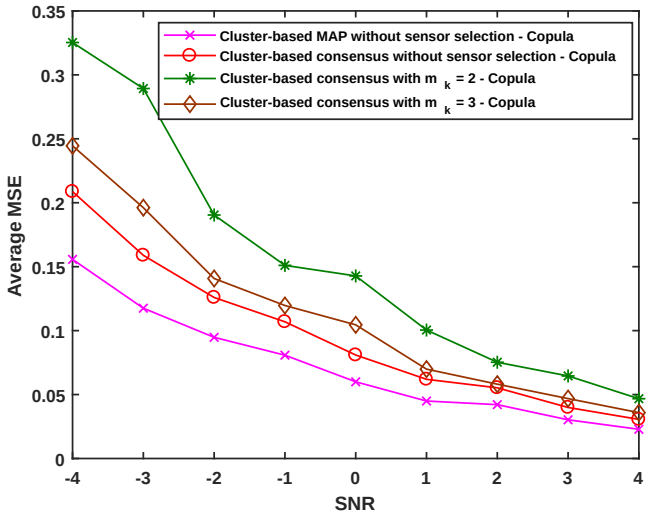

Fig. 6.8: Average MSE as a function of SNR for different schemes without sensor selection.

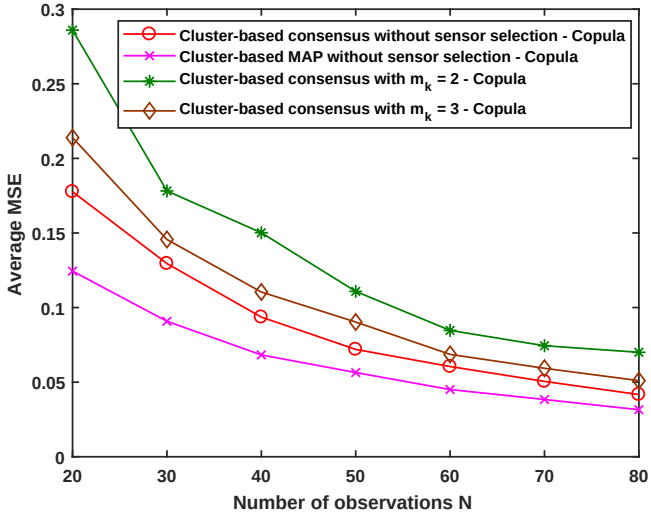

Fig. 6.9: Average MSE as a function of the number of observations N for different schemes without sensor selection.

sensus with $m_k = 2$ using copula based scheme and Cluster-based consensus with $m_k = 3$ using copula based scheme. As we can see that, selecting $m_k = 3$ sensors in each cluster results in better estimation performance compared to selecting $m_k = 2$ sensors in each cluster. Moreover, in Fig. 6.8, our proposed cluster-based consensus approach by selecting $m_k = 3$ sensors in each cluster performs very close to the corresponding scheme without sensor selection for SNR from 1 dB to 4 dB. Also, we have similar performance in Fig. 6.9 when the

number of observations is larger than or equal to 60. For the SNR value smaller than 1 dB and the number of observations smaller than 60, the performance difference between the cluster-based consensus scheme by selecting $m_k = 3$ sensors and the cluster-based consensus scheme without sensor selection is large. This is due to the fact that with low SNR values or small number of observations, the estimate obtained from each cluster is relatively poor. However, for the corresponding scheme without sensor selection, it includes more sensors and contains more information.

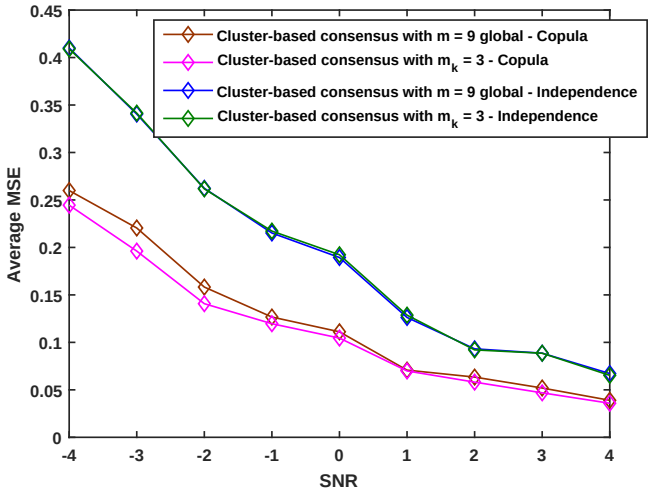

Fig. 6.10: Average MSE as a function of SNR for the cluster-based sensor selection scheme and the global sensor selection scheme.

In Fig. 6.10 and Fig. 6.11, we present the average MSE as a function of SNR and the number of observations N, respectively, for the cluster-based sensor selection scheme and the global sensor selection scheme (see Equation (6.17)). We evaluate the following schemes: Cluster-based consensus scheme using global sensor selection with $m = 9$ and copula based approach, Cluster-based consensus scheme using cluster-based selection scheme with $m_k = 3$ and copula based approach, Cluster-based consensus scheme using global sensor selection with $m = 9$ under independence assumption and Cluster-based consensus scheme using cluster-based selection scheme with $m_k = 3$ under independence assumption. Note that for fair comparison of the cluster-based sensor selection scheme and the global sensor selection scheme, we set

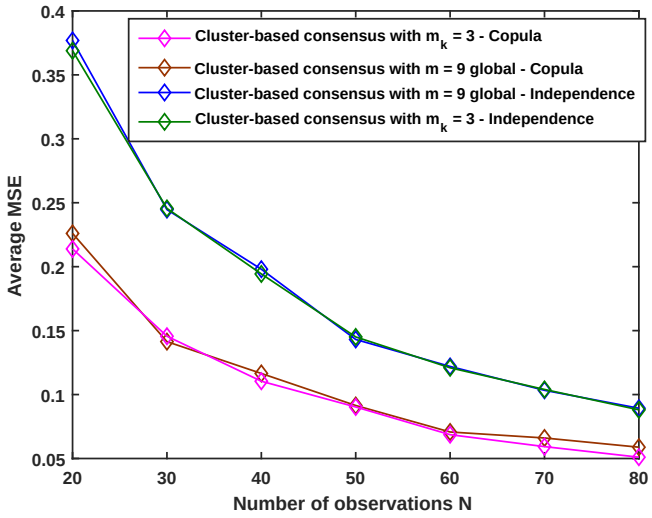

Fig. 6.11: Average MSE as a function of the number of observations N for the cluster-based sensor selection scheme and the global sensor selection scheme.

$m = 9$ as the total number of sensors that are selected since $m_k = 3$ sensors are selected from each cluster and the estimated number of clusters is $\hat{K} = 3$. As we can see, the cluster-based sensor selection scheme and global sensor selection scheme perform equally well.

6.6 Summary

In this chapter, a two-step cluster-based collaborative distributed estimation scheme was presented, where in the first step, sensors first form dependence driven clusters, and then perform copula-based MAP estimation via intra-cluster collaboration; in the second step, the estimates generated in the first step are shared via inter-cluster collaboration until an average consensus is reached. We proposed a merge based K-medoid dependence driven clustering algorithm. We further proposed a cluster-based sensor selection incorporated collaborative distributed estimation scheme. More specifically, prior to estimation, each cluster employs a mutual information based sensor selection scheme and selects sensors with maximum relevance and minimum redundancy with respect to the target parameter. Also, the proposed cluster-based sensor selection scheme was shown to be equivalent to the global/non-cluster based selection scheme with

high probability, and was computationally more efficient. Numerical results demonstrated the efficiency of our proposed methods compared to the estimation schemes under independence assumption.

CHAPTER 7
SUMMARY AND FUTURE DIRECTIONS

7.1 Summary

In this thesis, we investigated several inference problems with heterogeneous sources whose observations are statistically dependent. Both centralized and distributed inference problems were considered. By characterizing the statistical dependence among multimodal data, we proposed several methodologies that take such dependence into account to improve the inference performance. Also, we addressed the computational complexity issue resulting from modeling the underlying statistical dependence.

The problem of distributed detection with dependent sensor decisions was studied. We proposed a novel and powerful methodology to fuse dependent decisions obtained by binary quantization of statistically dependent sensor observations under the Neyman-Pearson framework. To derive the optimal fusion rule, we used the R-Vine copula model to characterize the complex dependence among multiple sensors. The proposed R-Vine copula based fusion methodology was employed to overcome the limitation of the existing standard multivariate copulas, and since this methodology is extremely flexible to model complex dependence structures. The optimal log likelihood test statistics at the FC involves multi-dimensional integration

at each time, leading to very high computational complexity. We proposed an efficient R-Vine copula based optimal fusion algorithm. Numerical results have illustrated the efficiency of our approach.

To tackle inference accuracy and response time issues for fusion of heterogeneous stream data, we proposed C-Storm, which is a novel parallel platform that is built based on Apache Storm and uses the copula-based dependence modeling approach for fusion of heterogeneous data streams. C-Storm offers fast inference responses and high inference accuracies. Moreover, it is a general and easy-to-use platform that can support various data fusion applications. Its users do not need to know details of Storm or copula theory. We demonstrated the superiority of C-Storm via a detection application. Experimental results have shown that C-Storm achieves 4.7x speedup over a sequential baseline on average, and higher degree of parallelism leads to better performance.

The problem of distributed classification for activity recognition with dependent high-level features was investigated. An R-Vine copula based feature fusion approach was presented to perform activity recognition using multi-modal sensor observations. The features of each modality were extracted via a DNN and afterwards, an R-Vine copula model was constructed to capture the dependencies of intra-modal and cross-modal features. The procedures of model construction involve selecting the optimal R-Vine tree structure, obtaining the copula parameter set, and choosing the best copula. Experiments on two human activity datasets demonstrated the efficiency of our proposed method compared to neural network based data/feature fusion, in terms of high prediction accuracy, less number of training samples required and dependence interpretability.

A distributed sequential detection problem was considered in a sensor network with spatially dependent observations. In this work, we proposed a copula-based sequential scheme for the problem of distributed detection with imperfect communication channels from the sensors to the fusion center. The underlying spatial dependence may result from the dependent

sensor signals, dependent channel noises or both. We used the regular vine copula model to represent the underlying dependence. We first proposed a centralized copula-based SPRT, and showed its asymptotic optimality and time efficiency. We then proposed a distributed copula-based sequential scheme, where the memory-less truncated SPRTs were performed at the local sensors and the copula-based SPRT was conducted at the FC. We have shown that by suitably designing the local thresholds and the truncation window, the local probability of false alarm and the local probability of miss detection of the proposed memoryless truncated local sequential tests are upper bounded by the pre-specified error probabilities. Moreover, we have shown the asymptotic optimality and time efficiency of the distributed copula-based SPRT. Via simulations, we have shown that our proposed copula-based SPRTs can efficiently capture the unknown dependence, and outperform the product-based SPRTs which ignore the underlying dependence.

Finally, we investigated a two-step cluster-based collaborative distributed estimation problem in a large scale wireless sensor network with dependent observations. In the first step, sensors first form dependence driven non-overlapping clusters, and then estimate the target parameter using copula based MAP via intra-cluster collaboration. In the second step, the estimates generated in the first step were shared via inter-cluster collaboration until an average consensus is reached. We proposed a merge based K-medoid dependence driven clustering algorithm. Moreover, we further proposed a cluster-based sensor selection incorporated collaborative distributed estimation scheme. More specifically, prior to the estimation, each cluster performs a mutual information based selection scheme by selecting sensors with maximum relevance and minimum redundancy with respect to the target parameter. Also, the proposed cluster-based sensor selection scheme is shown to be equivalent to the global/non-cluster based selection scheme with high probability, which at the same time is computationally more efficient. Numerical results demonstrated the efficiency of our proposed methods compared to independence assumed estimation schemes.

7.2 Future Directions

Some promising directions for future work are listed in the following:

1. In Chapter 2, we have proposed the regular vine copula model for the fusion of binary local decisions with two underlying hypotheses. It is optimal in terms of the fusion scheme at the FC. At the local sensors, suboptimal binary thresholding detectors are used. Although it may not be feasible to design optimal local decision rules, as a future work, one can investigate multi-bit quantizers at the local sensors to further improve the detection performance. Also, it is of interest to study multiple hypotheses testing problem under dependent observations. Copula selection approaches for the specific task of multiple hypotheses testing is worth pursuing as a future work.

2. In Chapter 3, we have designed a parallel computing platform based on multivariate copula dependence modeling approach. Typically, multivariate copula functions are limited to two sensor case. It is of great necessity to design parallel processing inference platform using regular vine copula based approach which is more powerful in characterizing high dimensional dependence structures.

3. In Chapter 4, we designed the feature-level fusion rule using regular vine copula based approach for multi-sensor data. For more complex data, such as video or audio data, the extracted features using deep networks are of potentially very high dimension. The regular vine copula may not be capable of characterizing very high dimensional data. Note that these extracted high-dimensional features are not necessarily all dependent in practice. Motivated by this, sparsity constrained regular vine copula can be pursued as a future work to extend the capability of regular vine copula based approach. Also, this idea can be applied to the large scale distributed estimation problem in Chapter 6.

4. In Chapter 5, we proposed a synchronized copula based distributed sequential test. The FC needs to wait until all sensor decisions are received before performing the sequential

test which can be time consuming. As a future work, one can investigate asynchronized local sequential tests to further improve the detection time.

APPENDICES

A $A_{\mathbf{u}^t}$ in Log Test Statictics (2.6)

First, we define $\tilde{I} = \{ln | u_{ln} = 1, l = 1, 2, \ldots, L, n = 1, 2, \ldots, N\}$ as the index set with decisions 1. Note that $\tilde{I} \cup \tilde{I}^c = I$. Moreover, let $\tilde{I}_j$ be the subset of $\tilde{I}$ and the cardinality of the set $\tilde{I}_j$ is j. We further define $\tilde{I}_{jt} = \{i_1 n, i_2 n, \ldots, i_t n, 0 \leq t \leq j\}$ as the subset of $\tilde{I}_j$ where $t = \left|\tilde{I}_{jt}\right|$. If t is a even number, $\tilde{I}_{jt} = \tilde{I}^e_{jt}$, otherwise, $\tilde{I}_{jt} = \tilde{I}^o_{jt}$. Under hypothesis H_1, let $P_{\tilde{I}_{jt}(i_1 n, i_2 n, \ldots, i_t n)}$ denote the PMFs that only the decisions of $(i_1 n, i_2 n, \ldots, i_t n)$th sensors are 1's and that of the rest of sensors are 0's. Note $t = 0$ implies that no sensor makes a decision 1 and $P_{\tilde{I}_{j0}}$ is used to denote all 0 sensor decisions. Similarly, let $Q_{\tilde{I}_{jt}(i_1 n, i_2 n, \ldots, i_t n)}$ denote the PMFs under H_0.

In the following, we illustrate the process of obtaining $A_{\mathbf{u}^t} = \log \frac{\prod_{0 \leq k \leq t} \mathcal{P}_{\tilde{I}^e_{tk}}^{(-1)^t} \prod_{0 \leq k \leq t} \mathcal{Q}_{\tilde{I}^o_{tk}}^{(-1)^t}}{\prod_{0 \leq k \leq t} \mathcal{Q}_{\tilde{I}^e_{tk}}^{(-1)^t} \prod_{0 \leq k \leq t} \mathcal{P}_{\tilde{I}^o_{tk}}^{(-1)^t}}$, where $\mathcal{P}_{\tilde{I}_{tk}} = \prod_{\{i_1 n, i_2 n, \ldots, i_k n\} \in \tilde{I}_t} P_{\tilde{I}_{tk}(i_1 n, i_2 n, \ldots, i_k n)}$ and $\mathcal{Q}_{\tilde{I}_{tk}} = \prod_{\{i_1 n, i_2 n, \ldots, i_k n\} \in \tilde{I}_t} Q_{\tilde{I}_{tk}(i_1 n, i_2 n, \ldots, i_k n)}$.

First, for $t = 1$, we have $k = 0, 1$. $A_{\mathbf{u}^1}$ is given as $A_{\mathbf{u}^1} = \log \frac{Q_{\tilde{I}_{10}} P_{\tilde{I}_{11}(i_1 n)}}{P_{\tilde{I}_{10}} Q_{\tilde{I}_{11}(i_1 n)}}, 1 \leq i_1 \leq L$ which satisfies the $A_{\mathbf{u}^t}$ with $t = 1$.

For $t = 2$, we have $k = 0, 1, 2$. $A_{\mathbf{u}^2}$ is given as $A_{\mathbf{u}^2} = \log \frac{P_{\tilde{I}_{20}} P_{\tilde{I}_{22}(i_1 n, i_2 n)} Q_{\tilde{I}_{21}(i_1 n)} Q_{\tilde{I}_{21}(i_2 n)}}{Q_{\tilde{I}_{20}} Q_{\tilde{I}_{22}(i_1 n, i_2 n)} P_{\tilde{I}_{21}(i_1 n)} P_{\tilde{I}_{21}(i_2 n)}}, 1 \leq i_1, i_2 \leq L$ which satisfies $A_{\mathbf{u}^t}$ with $t = 2$.

For $t = 3$, we have $k = 0, 1, 2, 3$. $A_{\mathbf{u}^3}$ is given as $A_{\mathbf{u}^3} = \log \frac{\mathcal{Q}_{\tilde{I}^e_{30}} \mathcal{Q}_{\tilde{I}^e_{32}} \mathcal{P}_{\tilde{I}^o_{31}} \mathcal{P}_{\tilde{I}^o_{33}}}{\mathcal{P}_{\tilde{I}^e_{30}} \mathcal{P}_{\tilde{I}^e_{32}} \mathcal{Q}_{\tilde{I}^o_{31}} \mathcal{Q}_{\tilde{I}^o_{33}}}, 1 \leq i_1, i_2, i_3 \leq L$, where for the numerator, $\mathcal{Q}_{\tilde{I}^e_{30}} = Q_{\tilde{I}_{30}}$, $\mathcal{Q}_{\tilde{I}^e_{32}} = Q_{\tilde{I}_{32}(i_1 n, i_2 n)} Q_{\tilde{I}_{32}(i_1 n, i_3 n)} Q_{\tilde{I}_{32}(i_2 n, i_3 n)}$, $\mathcal{P}_{\tilde{I}^o_{31}} = P_{\tilde{I}_{31}(i_1 n)} P_{\tilde{I}_{31}(i_2 n)} P_{\tilde{I}_{31}(i_3 n)}$ and $\mathcal{P}_{\tilde{I}^o_{33}} = P_{\tilde{I}_{33}(i_1 n, i_2 n, i_3 n)}$. We can verify that $A_{\mathbf{u}^3}$ satisfies $A_{\mathbf{u}^t}$ with $t = 3$.

For $t = 4$, we have $k = 0, 1, 2, 3, 4$. $A_{\mathbf{u}^4}$ is given as $A_{\mathbf{u}^4} = \log \frac{\mathcal{P}_{\tilde{I}^e_{40}} \mathcal{P}_{\tilde{I}^e_{42}} \mathcal{P}_{\tilde{I}^e_{44}} \mathcal{Q}_{\tilde{I}^o_{41}} \mathcal{Q}_{\tilde{I}^o_{43}}}{\mathcal{Q}_{\tilde{I}^e_{40}} \mathcal{Q}_{\tilde{I}^e_{42}} \mathcal{Q}_{\tilde{I}^e_{44}} \mathcal{P}_{\tilde{I}^o_{41}} \mathcal{P}_{\tilde{I}^o_{43}}}, 1 \leq$

$i_1, i_2, i_3, i_4 \leq L$ which satisfies $A_{\mathbf{u}^t}$ with $t = 4$.

For $t = 5, 6, \ldots, L$, $A_{\mathbf{u}^t}$ can be easily verified. ■

B Proof of Theorem 2.1

Note that Equation (2.6) can be written as

$$\log\Lambda(\mathbf{u}) = \sum_{n=1}^{N} U_n, \tag{8.1}$$

where $U_n = \sum_{\{i_1 n\} \in I_1} A_{\mathbf{u}^1} \mathbf{u}^1 + \sum_{\{i_1 n, i_2 n\} \in I_2} A_{\mathbf{u}^2} \mathbf{u}^2 + \ldots + \sum_{\{i_1 n, i_2 n, \ldots, i_t n\} \in I_t} A_{\mathbf{u}^t} \mathbf{u}^t + \ldots + \sum_{\{i_1 n, i_2 n, \ldots, i_L n\} \in I_L} A_{\mathbf{u}^L} \mathbf{u}^L$.

Due to the assumption of temporal independence of sensor decisions, U_n for all $1 \leq n \leq N$ are i.i.d. random variables. Hence, by applying the central limit theorem (CLT) [82], $\log\Lambda(\mathbf{u})$ is asymptotically Gaussian.

Note that $u_{ln}, l = 1, 2, \ldots, L$ are Bernoulli distributed under both hypotheses and can take a value of either 0 or 1 with certain probabilities. For the simplification of notation, we omit the time index n here. For sensor decisions $s \in S$, we define $E = \{j_1, j_2, \ldots, j_d, 1 \leq d \leq L\}$ as the index set when the sensor decisions of s are 1. Under H_1 hypothesis, the random variable $U_s = \sum_{\{j_{m_1}\} \subset E} A_{\mathbf{u}^1} + \sum_{\{j_{m_1}, j_{m_2}\} \subset E} A_{\mathbf{u}^2} + \ldots + \sum_{\{j_{m_1}, j_{m_2}, \ldots, j_{m_{d-1}}\} \subset E} A_{\mathbf{u}^{d-1}} + \sum_{\{j_{m_1}, j_{m_2}, \ldots, j_{m_d}\} \subset E} A_{\mathbf{u}^d}$ with probability P_s for $1 \leq d \leq L$, otherwise, $U = 0$ for $d = 0$. Similarly, we can obtain the values of U under H_0 hypothesis. Since we can obtain the joint PMF of sensor decisions by integrating the joint PDF of their observations under both hypotheses, we now can evaluate the mean and variance of the Gaussian distributed fusion statistic under either hypothesis. The mean and variance of the fusion rule statistic under both hypotheses are

given as follows

$$\mu_0 = N\left[\sum_{s\in S} U_s Q_s\right],$$
$$\sigma_0^2 = N\left[\sum_{s\in S} U_s^2 Q_s - (\mu_0/N)^2\right],$$
$$\mu_1 = N\left[\sum_{s\in S} U_s P_s\right],$$
$$\sigma_1^2 = N\left[\sum_{s\in S} U_s^2 P_s - (\mu_1/N)^2\right].$$

■

C Proof of Theorem 5.1

For notational simplicity, we remove the superscript “cen" in the proof. Before deriving the expected stopping time, we first present Wald’s Identity [99] which is given by the following Proposition.

Proposition 8.1 (Wald’s Identity)**.** *Let $Y_1, Y_2, \ldots$ be independent and identically distributed random variables with mean μ. Let K be any integer-valued random variable such that $\mathbb{E}[K] < \infty$ and $K = k$ is an event determined by $Y_1, Y_2, \ldots, Y_k$ and independent of $Y_i, i > k$. Then $\mathbb{E}[\sum_{i=1}^{K} Y_i] = \mu\,\mathbb{E}[K]$.*

Let $\ell(i)$ denote the log-likelihood ratio at time instant i, which is given as

$$\ell(i) = \log \frac{f_1(\mathbf{y}_i)}{f_0(\mathbf{y}_i)}, \tag{8.2}$$

where $f_k(\mathbf{y}_i), k = 0, 1$ are the joint distributions of observations $[y_{1i}, \ldots, y_{Li}]$ under the two hypotheses. Moreover, the log-likelihood ratio test statistic $\Lambda^n(\mathbf{y})$ is given as

$$\Lambda^n(\mathbf{y}) = \sum_{i=1}^{n} \ell(i), \tag{8.3}$$

where $n = 1, 2, \ldots$ and we assume that $\Lambda^0(\mathbf{y}) = 0$.

Firstly, we show that as $A \to \infty$ and $B \to \infty$, $P_F \approx e^{-A}$ and $P_M \approx e^{-B}$. Under H_0 hypothesis, based on Equation (8.2) and Equation (8.3), we consider

$$\begin{aligned}
&\mathbb{E}_{H_0}\left[e^{\Lambda^{n+1}(\mathbf{y})}|e^{\Lambda^1(\mathbf{y})}, \ldots, e^{\Lambda^n(\mathbf{y})}\right] \\
&= \mathbb{E}_{H_0}\left[\frac{f_1(\mathbf{y}_{n+1})}{f_0(\mathbf{y}_{n+1})}e^{\Lambda^n(\mathbf{y})}|e^{\Lambda^1(\mathbf{y})}, \ldots, e^{\Lambda^n(\mathbf{y})}\right], \\
&= e^{\Lambda^n(\mathbf{y})}\mathbb{E}_{H_0}\left[\frac{f_1(\mathbf{y}_{n+1})}{f_0(\mathbf{y}_{n+1})}|e^{\Lambda^1(\mathbf{y})}, \ldots, e^{\Lambda^n(\mathbf{y})}\right], \\
&= e^{\Lambda^n(\mathbf{y})}\mathbb{E}_{H_0}\left[\frac{f_1(\mathbf{y}_{n+1})}{f_0(\mathbf{y}_{n+1})}\right], \\
&\overset{(a)}{=} e^{\Lambda^n(\mathbf{y})},
\end{aligned}$$

where (a) is obtained since $\mathbf{y}_1, \mathbf{y}_2, \ldots$ are independent, we have

$$\mathbb{E}_{H_0}\left\{e^{\ell(i)}\right\} = \mathbb{E}_{H_0}\left[\frac{f_1(\mathbf{y}_i)}{f_0(\mathbf{y}_i)}\right] = \int_{\mathcal{R}} \frac{f_1(\mathbf{y}_i)}{f_0(\mathbf{y}_i)} f_0(\mathbf{y}_i) d\mathbf{y}_i = 1. \tag{8.4}$$

Therefore, the $e^{\Lambda^n(\mathbf{y})}$ forms a martingale under H_0 hypothesis. When the test stops, we have

$$\Lambda^N(\mathbf{y}) = A \quad \text{or} \quad \Lambda^N(\mathbf{y}) = -B. \tag{8.5}$$

According to Wald's likelihood ratio identity [99, Theorem 2.3.3], we have

$$\mathbb{E}_{H_0}[e^{\Lambda^N(\mathbf{y})}] = 1. \tag{8.6}$$

Combining Equation (8.5) and Equation (8.6), we have

$$1 = \mathbb{E}_{H_0}[e^{\Lambda^N(\mathbf{y})}] \approx P_F e^A + (1 - P_F)e^{-B}. \tag{8.7}$$

Therefore, we have

$$P_F \approx \frac{1-e^{-B}}{e^{A}-e^{-B}}. \tag{8.8}$$

Similarly, we can show that $e^{-\Lambda^n(\mathbf{y})}$ forms a martingale under H_1 hypothesis and obtain

$$P_M \approx \frac{e^{-A}-1}{e^{-A}-e^{B}}. \tag{8.9}$$

As $A, B \to \infty$, the P_F and P_M in Equation (8.8) and Equation (8.9), respectively, are reduced to

$$P_F \approx e^{-A}, \quad P_M \approx e^{-B}. \tag{8.10}$$

Secondly, we derive the asymptotic expected stopping time $\mathbb{E}_k[N]$ under hypothesis $H_k, k = 0, 1$. Under hypothesis H_0, by Wald's Identity,

$$\begin{aligned}
&\mathbb{E}_{H_0}\left[\Lambda^N(\mathbf{y})\right] \\
&= \mathbb{E}_{H_0}\left[\sum_{i=1}^{N}\sum_{l=1}^{L}\log\frac{f_{1,l}(y_{li})}{f_{0,l}(y_{li})} + \sum_{i=1}^{N}\log\frac{c_1(\cdot|\boldsymbol{\phi}_1)}{c_0(\cdot|\boldsymbol{\phi}_0)}\right], \\
&= \mathbb{E}_{H_0}[N] * \mathbb{E}_{H_0}\left[\sum_{l=1}^{L}\log\frac{f_{1,l}(y_{li})}{f_{0,l}(y_{li})} + \log\frac{c_1(\cdot|\boldsymbol{\phi}_1)}{c_0(\cdot|\boldsymbol{\phi}_0)}\right], \\
&\overset{(b)}{=} \mathbb{E}_{H_0}[N] * \\
&\left[-\sum_{l=1}^{L} D\left(f_{0,l}(\cdot)||f_{1,l}(\cdot)\right) - D\left(c_0(\cdot|\boldsymbol{\phi}_0)||c_1(\cdot|\boldsymbol{\phi}_1)\right)\right],
\end{aligned} \tag{8.11}$$

where (b) is obtained based on the fact that copula-based dependence measure is independent of the marginal distributions.

Moreover, we have

$$\begin{aligned}
\mathbb{E}_{H_0}\left[\Lambda^N(\mathbf{y})\right] &\approx P_F * A + (1-P_F) * (-B), \\
&\overset{(c)}{\approx} -B,
\end{aligned} \tag{8.12}$$

where (c) is obtained by using Equation (8.10).

By combining Equation (8.11) and Equation (8.12), we obtain $\mathbb{E}_{H_0}[N]$ in Equation (5.6). Similarly, we can get $\mathbb{E}_{H_1}[N]$. ■

D Proof of Theorem 5.2

Before we proceed, we define the CDF of $\lambda_{l,i}(iW_0)$ under hypothesis $H_k, k = 0, 1$ as $\tilde{Q}_k(x) = P(\lambda_{l,i}(iW_0) \leq x|H_k)$.

We first derive the upper bound of $P_F^{\text{local}} = P(U_{li} = 1|H_0)$. Here, we define $\mathcal{V}$ as $P(\nu_l \leq \lambda_{l,i}(iW_0) < a_l, -b_l < \lambda_{l,i}(W) < a_l, (i-1)W_0 + 1 \leq W \leq iW_0)$. Therefore, P_F^{local} is given by

$$\begin{aligned} P_F^{\text{local}} &< \sum_{W=(i-1)W_0+1}^{iW_0} P(U_{li}^W = 1|H_0) + \mathcal{V}, \\ &< 1 - Q_0(a_l) + \mathcal{V}, \\ &\overset{(a)}{<} \frac{\tilde{\alpha}}{\Delta_{l,a}} + 1 - \tilde{Q}_0(\nu_l), \end{aligned} \tag{8.13}$$

where (a) is obtained by using the upper bound derived in Appendix [98, A.4].

If we set W_0 equal to N_{FSS} and ν_l equal to ν_{FSS}, we have $1 - Q_0(\nu_l) = \frac{\tilde{\alpha}}{\Delta_{l,Ta}}$, and the inequality in Equation (8.13) can be written as

$$P_F^{\text{local}} < \left(\frac{1}{\Delta_{l,a}} + \frac{1}{\Delta_{l,Ta}}\right)\tilde{\alpha}. \tag{8.14}$$

Then, we derive the upper bound of $P_M^{\text{local}} = P(U_{li} = 0|H_1)$. Similarly, we define $\mathcal{U} = P(-b_l < \lambda_{l,i}(iW_0) < \nu_l, -b_l < \lambda_{l,i}(W) < a_l, (i-1)W_0 + 1 \leq W \leq iW_0|H_1)$. P_M^{local} can be

written as

$$\begin{aligned} P_M^{\text{local}} &< \sum_{W=(i-1)W_0+1}^{iW_0} P(U_{li}^W = 0|H_1) + \mathcal{U}, \\ &< \frac{\tilde{\beta}}{\Delta_{l,b}} + \mathcal{U}, \\ &\overset{(b)}{<} \frac{\tilde{\beta}}{\Delta_{l,b}} + \tilde{Q}_1(\nu_l), \end{aligned} \tag{8.15}$$

where (b) is obtained by using the upper bound derived in Appendix [98, A.3]. By setting W_0 equal to N_{FSS} and ν_l equal to ν_{FSS}, we can show that $P_M^{\text{local}} < \left(\frac{1}{\Delta_{l,b}} + \frac{1}{\Delta_{l,Tb}}\right)\tilde{\beta}$. ■

E Proof of Theorem 5.3

Noting that $T_p = N_{T_p} * \min\{\mathbb{E}[t_{\text{max}}], W_0\}$ and by invoking Wald's Identity, we have

$$\begin{aligned} &\mathbb{E}_k[T_p] = \mathbb{E}_k[\Lambda^{T_p}] * \min\{\mathbb{E}[t_{\text{max}}], W_0\} \\ &/ \left(\sum_{l=1}^{L} \mathbb{E}_k\left[\log \frac{f_{1,l}(\cdot)}{f_{0,l}(\cdot)}\right] + \mathbb{E}_k\left[\log \frac{c_1(\cdot|\boldsymbol{\phi}_1)}{c_0(\cdot|\boldsymbol{\phi}_0)}\right]\right). \end{aligned} \tag{8.16}$$

Similar to the proof of Theorem 5.1, we have $P_F \approx e^{-A}$ and $P_M \approx e^{-B}$. As $B \to \infty$, we have

$$\mathbb{E}_1[T_p] = \frac{A * \min\{\mathbb{E}[t_{\text{max}}], W_0\}}{D_1}, \tag{8.17}$$

where $D_1 = \sum_{l=1}^{L} D(f_{1,l}(\cdot)||f_{0,l}(\cdot)) + D(c_1(\cdot|\boldsymbol{\phi}_1)||c_0(\cdot|\boldsymbol{\phi}_0))$.

Since $\mathbb{E}[t_{\text{max}}] < W_0$, in the worse case, we obtain Equation (5.21). Similarly, under hypothesis H_0, we can obtain Equation (5.20). ■

F Proof of Theorem 6.1

$$\begin{aligned}
\text{var}\left(\hat{\theta}\right) &= \text{var}\left(\frac{1}{K}\sum_{k=1}^{K}\hat{\theta}_k\right), \\
&= \frac{1}{K^2}\left[\sum_{k=1}^{K}\text{var}\left(\hat{\theta}_k\right) + \sum_{k=1}^{K}\sum_{\tilde{k}\neq k=1}^{K}\text{cov}\left(\hat{\theta}_k, \hat{\theta}_{\tilde{k}}\right)\right], \\
&\overset{(b)}{\leq} \frac{1}{K^2}\left[\sum_{k=1}^{K}\text{var}\left(\hat{\theta}_k\right) + \sum_{k=1}^{K}\sum_{\tilde{k}\neq k=1}^{K}\sqrt{\text{var}\left(\hat{\theta}_k\right)\text{var}\left(\hat{\theta}_{\tilde{k}}\right)}\right], \\
&= \left(\frac{1}{K}\sum_{k=1}^{K}\sqrt{\text{var}\left(\hat{\theta}_k\right)}\right)^2,
\end{aligned}$$

where (b) is obtained using $\text{cov}\left(\hat{\theta}_k, \hat{\theta}_{\tilde{k}}\right) / \sqrt{\text{var}\left(\hat{\theta}_k\right)\text{var}\left(\hat{\theta}_{\tilde{k}}\right)} \leq 1$. Thus, we obtain that

$$\sqrt{\text{var}\left(\hat{\theta}\right)} \leq \frac{1}{K}\sum_{k=1}^{K}\sqrt{\text{var}\left(\hat{\theta}_k\right)}.$$

■

G Proof of Theorem 6.2

Suppose that we already have the set $\mathbf{s}$ which consists of selected sensors using the global incremental sensor selection scheme in (6.17). We can trace back these selected sensors in the set $\mathbf{s}$ to clusters. Without loss of generality, we assume that the sensors in the set $\mathbf{s}$ belong to clusters $\mathcal{G}_1, \ldots, \mathcal{G}_i, i < K$, and we decompose the set $\mathbf{s}$ into $\mathbf{s}_1, \ldots, \mathbf{s}_i$ with $s_r, r = 1, \ldots, i$ denoting the subset of sensors that belongs to cluster $\mathcal{G}_r$.

Assume that we have a candidate data sequence $\mathbf{z}_{\tilde{j}} \in \mathbf{z}_{\mathcal{S}} \setminus \mathbf{z}_{\mathbf{s}}$, where $\mathbf{z}_{\mathcal{S}}$ is the set of data sequences obtained from all the sensors in the network. Therefore, the global incremental

selection problem becomes:

$$\max_{\mathbf{z}_{\tilde{j}} \in \mathbf{z}_{\mathcal{S}} \setminus \mathbf{z}_{\mathbf{s}}} I(\mathbf{z}_{\tilde{j}}; \theta) - \frac{1}{|\mathbf{s}|} \sum_{\mathbf{z}_t \in \mathbf{z}_{\mathbf{s}}} I(\mathbf{z}_{\tilde{j}}; \mathbf{z}_t). \tag{8.18}$$

Note that there are two cases for the assignment of the sequence $\mathbf{z}_{\tilde{j}}$. The first case is that $\mathbf{z}_{\tilde{j}}$ belongs to one of the clusters in set $[\mathcal{G}_1, \ldots, \mathcal{G}_i]$. The second case is that $\mathbf{z}_{\tilde{j}}$ belongs to one of the clusters in set $[\mathcal{G}_{i+1}, \ldots, \mathcal{G}_K]$.

For the first case, without loss of generality, we assume that $\mathbf{z}_{\tilde{j}}$ belongs to cluster $\mathcal{G}_j$. Also, we further suppose that set $\mathbf{s}_j \in \mathbf{s}$ contains the selected sensors from cluster $\mathcal{G}_j, j \leq i$. Thus, the problem in Equation (8.18) can be further decomposed into the following problem:

$$\begin{aligned} \max_{\mathbf{z}_{\tilde{j}} \in \mathbf{z}_{\mathcal{G}_j} \setminus \mathbf{z}_{\mathbf{s}_j}} I(\mathbf{z}_{\tilde{j}}; \theta) - \frac{1}{|\mathbf{s}|} \sum_{\mathbf{z}_{\tilde{t}} \in \mathbf{z}_{\mathbf{s}}} I(\mathbf{z}_{\tilde{j}}; \mathbf{z}_{\tilde{t}}) \\ - \frac{1}{|\mathbf{s}|} \sum_{\mathbf{z}_t \in \mathbf{z}_{\mathbf{s}} \setminus \mathbf{z}_{\mathbf{s}_j}} I(\mathbf{z}_{\tilde{j}}; \mathbf{z}_t). \end{aligned} \tag{8.19}$$

For the second case, the problem in Equation (8.18) becomes

$$\max_{\mathbf{z}_{\tilde{j}} \in \cup_{\tilde{i}=i+1}^{K} \mathbf{z}_{\mathcal{G}_{\tilde{i}}}} I(\mathbf{z}_{\tilde{j}}; \theta) - \frac{1}{|\mathbf{s}|} \sum_{\mathbf{z}_t \in \mathbf{z}_{\mathbf{s}}} I(\mathbf{z}_{\tilde{j}}; \mathbf{z}_t) \tag{8.20}$$

Note that for the problems in Equation (8.19) and Equation (8.20), we have $\mathbf{z}_t$ and $\mathbf{z}_{\tilde{j}}$ that are generated from different clusters. Using Assumption A_2, we have $P\left(d(\mathbf{z}_{\tilde{j}}, \mathbf{z}_t) > d_{th}\right) \geq 1 - \epsilon$, where $d_{th}, d_L < d_{th} < d_H$ is the threshold we used to cluster sensors. If the dissimilarity of two data sequences is greater than d_{th}, we put these sequences into two into different clusters; Otherwise, we put them into the same cluster. $\epsilon > 0$ is a small allowed tolerance. Furthermore, we assume that $I(\mathbf{z}_{\tilde{j}}; \mathbf{z}_t)$ is a non-increasing function of the dissimilarity $d(\mathbf{z}_{\tilde{j}}, \mathbf{z}_t)$. Based on Assumption A_2, we have

$$P(I(\mathbf{z}_{\tilde{j}}; \mathbf{z}_t) < \zeta) \geq 1 - \epsilon, \tag{8.21}$$

where ζ is the obtained mutual information with dissimilarity d_{th}. Note that the empirical mutual information $I(\mathbf{z}_{\tilde{j}}; \mathbf{z}_t)$ also depends on the number of data samples that are available. In this proof, we assume that we have enough data samples to estimate the empirical mutual information accurately.

Note that the closed form expression for ζ is difficult to obtain due to the complicated relationship between the mutual information and rank-based dependence measure. The mutual information and the rank-based dependence measure (Spearman's ρ or Kendall's τ) can be connected using the copula based dependence measure. For random variables x and y, the connection between mutual information and copula-based dependence measure is given as

$$I(x; y) = \int_{[0,1]^2} c(u, v) \log c(u, v) d_u d_v, \tag{8.22}$$

where c is the copula density function between variables x and y. Also, $u = F(x)$ and $v = F(y)$, where $F(\cdot)$ is the CDF.

The connections between the rank-based dependence measures (Kendall's τ and Spearman's ρ) and the copula-based dependence measure are given as

$$\tau(x, y) = 4 \int_{[0,1]^2} C(u, v) dC(u, v) - 1,$$
$$\rho(x, y) = 12 \int_u \int_v C(u, v) du dv - 3.$$

The computation of ζ can be carried our using numerical differentiation and integration. However, if x and y follow Gaussian distributions and are linearly dependent, we have

$$I(x; y) = -\frac{1}{2} \log(1 - r^2), \tag{8.23}$$

where $r = \text{corr}(x, y)$ is the Pearson correlation coefficient. If we define our dissimilarity as $d(\cdot, \cdot) = \sqrt{1 - r^2}$, we have $r_{th}^2 = 1 - d_{th}^2$ given d_{th}. Therefore, $\zeta = -\log d_{th}$. As we can see

that, ζ is a decreasing function of d_{th}.

In the following, our goal is to show that for the first case $P\left(\frac{1}{|\mathbf{s}|}\sum_{\mathbf{z}_t\in\mathbf{z}_\mathbf{s}\setminus\mathbf{z}_{\mathbf{s}_j}} I(\mathbf{z}_{\tilde{j}};\mathbf{z}_t)>\zeta\right) < n_1\epsilon$, where $1 \leq n_1 < L$, and for the second case, $P\left(\frac{1}{|\mathbf{s}|}\sum_{\mathbf{z}_t\in\mathbf{z}_\mathbf{s}} I(\mathbf{z}_{\tilde{j}};\mathbf{z}_t)>\zeta\right) < n_2\epsilon$, where $1 \leq n_2 < L$.

We first prove for the first case.

$$
\begin{aligned}
&P\left(\frac{1}{|\mathbf{s}|}\sum_{\mathbf{z}_t\in\mathbf{z}_\mathbf{s}\setminus\mathbf{z}_{\mathbf{s}_j}} I(\mathbf{z}_{\tilde{j}};\mathbf{z}_t)>\zeta\right)\\
&< P\left(\frac{1}{|\mathbf{s}\setminus\mathbf{s}_j|}\sum_{\mathbf{z}_t\in\mathbf{z}_\mathbf{s}\setminus\mathbf{z}_{\mathbf{s}_j}} I(\mathbf{z}_{\tilde{j}};\mathbf{z}_t)>\zeta\right),\\
&< P\left(\max_{\mathbf{z}_t\in\mathbf{z}_\mathbf{s}\setminus\mathbf{z}_{\mathbf{s}_j}} I(\mathbf{z}_{\tilde{j}};\mathbf{z}_t)>\zeta\right),\\
&< \sum_{\mathbf{z}_t\in\mathbf{z}_\mathbf{s}\setminus\mathbf{z}_{\mathbf{s}_j}} P\left(I(\mathbf{z}_{\tilde{j}};\mathbf{z}_t)>\zeta\right),\\
&< |\mathbf{s}\setminus\mathbf{s}_j|\epsilon,
\end{aligned}
$$

where $1 \leq |\mathbf{s}\setminus\mathbf{s}_j| < L$.

Therefore, we have $P\left(\frac{1}{|\mathbf{s}|}\sum_{\mathbf{z}_t\in\mathbf{z}_\mathbf{s}\setminus\mathbf{z}_{\mathbf{s}_j}} I(\mathbf{z}_{\tilde{j}};\mathbf{z}_t)>\zeta\right) < n_1\epsilon$, where $1 \leq n_1 < L$. Similarly, we can show that $P\left(\frac{1}{|\mathbf{s}|}\sum_{\mathbf{z}_t\in\mathbf{z}_\mathbf{s}} I(\mathbf{z}_{\tilde{j}};\mathbf{z}_t)>\zeta\right) < n_2\epsilon$, where $1 \leq n_2 < L$.

By suitably designing d_{th}, we can make ζ sufficiently small. Therefore, with probability at least $1-\epsilon$, the term $\frac{1}{|\mathbf{s}|}\sum_{\mathbf{z}_t\in\mathbf{z}_\mathbf{s}\setminus\mathbf{z}_{\mathbf{s}_j}} I(\mathbf{z}_{\tilde{j}};\mathbf{z}_t)$ and the term $\frac{1}{|\mathbf{s}|}\sum_{\mathbf{z}_t\in\mathbf{z}_\mathbf{s}} I(\mathbf{z}_{\tilde{j}};\mathbf{z}_t)$ are upper bounded by ζ.

For the first case, by ignoring the term $\frac{1}{|\mathbf{s}|}\sum_{\mathbf{z}_t\in\mathbf{z}_\mathbf{s}\setminus\mathbf{z}_{\mathbf{s}_j}} I(\mathbf{z}_{\tilde{j}};\mathbf{z}_t)$, the problem in Equation (8.19) reduces to

$$\max_{\mathbf{z}_{\tilde{j}}\in\mathbf{z}_{\mathcal{G}_j}\setminus\mathbf{z}_{\mathbf{s}_j}} I(\mathbf{z}_{\tilde{j}};\theta) - \frac{1}{|\mathbf{s}|}\sum_{\mathbf{z}_{\tilde{t}}\in\mathbf{z}_{\mathbf{s}_j}} I(\mathbf{z}_{\tilde{j}};\mathbf{z}_{\tilde{t}}). \tag{8.24}$$

Since $\frac{1}{|\mathbf{s}|}$ is a scale parameter, which will not affect the solution of the problem in Equation

(8.24), the above optimization problem can be further written as

$$\max_{\mathbf{z}_{\tilde{j}} \in \mathbf{z}_{\mathcal{G}_j} \setminus \mathbf{z}_{\mathbf{s}_j}} I(\mathbf{z}_{\tilde{j}}; \theta) - \frac{1}{|\mathbf{s}_j|} \sum_{\mathbf{z}_{\tilde{t}} \in \mathbf{z}_{\mathbf{s}_j}} I(\mathbf{z}_{\tilde{j}}; \mathbf{z}_{\tilde{t}}), \tag{8.25}$$

which is equivalent to the cluster-based incremental search problem in Equation (6.18).

For the second case, by ignoring the term $\frac{1}{|\mathbf{s}|} \sum_{\mathbf{z}_t \in \mathbf{z}_\mathbf{s}} I(\mathbf{z}_{\tilde{j}}; \mathbf{z}_t)$, the problem in Equation (8.20) reduces to cluster-based incremental search problem in Equation (6.18) while selecting the first sensor in the cluster.

Remark 8.1. *Since we don't consider weakly dependent sensors (nearly independent sensors) within a cluster in this work, for the first case in Equation* (8.19)*, the term* $\frac{1}{|\mathbf{s}|} \sum_{\mathbf{z}_{\tilde{t}} \in \mathbf{z}_\mathbf{s}} I(\mathbf{z}_{\tilde{j}}; \mathbf{z}_{\tilde{t}})$ *is significantly larger than the term* $\frac{1}{|\mathbf{s}|} \sum_{\mathbf{z}_t \in \mathbf{z}_\mathbf{s} \setminus \mathbf{z}_{\mathbf{s}_j}} I(\mathbf{z}_{\tilde{j}}; \mathbf{z}_t)$*. The extreme scenario is that the term* $\frac{1}{|\mathbf{s}|} \sum_{\mathbf{z}_{\tilde{t}} \in \mathbf{z}_\mathbf{s}} I(\mathbf{z}_{\tilde{j}}; \mathbf{z}_{\tilde{t}})$ *is a small number due to a large scale parameter* $|\mathbf{s}|$*. For this scenario, the dominant term would be* $I(\mathbf{z}_{\tilde{j}}; \theta)$ *which can be covered by the second case in Equation* (8.20)*.*

Therefore, by designing d_{th}, with at least probability $1 - \epsilon$, the global incremental search method in Equation (6.17) reduces to cluster-based incremental search. ■

REFERENCES

[1] V. Aalo and R. Viswanathan, "On distributed detection with correlated sensors: Two examples," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, no. 3, pp. 414–421, 1989.

[2] K. Aas, C. Czado, A. Frigessi, and H. Bakken, "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and economics*, vol. 44, no. 2, pp. 182–198, 2009.

[3] H. Akaike, B. Petrov, and F. Csaki, "Information theory and an extension of the maximum likelihood principle," 1973.

[4] M. Anderson, T. Adali, and X.-L. Li, "Joint blind source separation with multivariate gaussian model: Algorithms and performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1672–1683, 2011.

[5] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones." in *Esann*, 2013.

[6] K. R. Apt and A. Witzel, "A generic approach to coalition formation," *International Game Theory Review*, vol. 11, no. 03, pp. 347–367, 2009.

[7] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Challenges and applications in multimodal machine learning," in *The Handbook of Multimodal-Multisensor Interfaces*. Association for Computing Machinery and Morgan & Claypool, 2018, pp. 17–48.

[8] ——, "Multimodal machine learning: A survey and taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2018.

[9] O. Banos, M. Damas, H. Pomares, F. Rojas, B. Delgado-Marquez, and O. Valenzuela, "Human activity recognition based on a sensor weighting hierarchical classifier," *Soft Computing*, vol. 17, no. 2, pp. 333–343, 2013.

[10] R. Bechhofer, "A note on the limiting relative efficiency of the wald sequential probability ratio test," *Journal of the American Statistical Association*, vol. 55, no. 292, pp. 660–663, 1960.

[11] T. Bedford and R. M. Cooke, "Probability density decomposition for conditionally dependent random variables modeled by vines," *Annals of Mathematics and Artificial intelligence*, vol. 32, no. 1-4, pp. 245–268, 2001.

[12] ——, "Vines: A new graphical model for dependent random variables," *Annals of Statistics*, pp. 1031–1068, 2002.

[13] S. Blackman and R. Popoli, *Design and analysis of modern tracking systems*. Norwood, MA: Artech House, 1999.

[14] R. Bramon, I. Boada, A. Bardera, J. Rodriguez, M. Feixas, J. Puig, and M. Sbert, "Multimodal data fusion based on mutual information," *IEEE transactions on visualization and computer graphics*, vol. 18, no. 9, pp. 1574–1587, 2011.

[15] T. Butz and J.-P. Thiran, "From error probability to information theoretic (multi-modal) signal processing," *Signal Processing*, vol. 85, no. 5, pp. 875–902, 2005.

[16] F. S. Cattivelli and A. H. Sayed, "Diffusion lms strategies for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, 2010.

[17] Z. Chair and P. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, no. 1, pp. 98–101, 1986.

[18] J.-F. Chamberland and V. V. Veeravalli, "How dense should a sensor network be for detection with correlated observations?" *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5099–5106, 2006.

[19] U. Cherubini, E. Luciano, and W. Vecchiato, *Copula Methods in Finance*. Prentice Hall, 2004.

[20] A. Chiuso, F. Fagnani, L. Schenato, and S. Zampieri, "Gossip algorithms for simultaneous distributed estimation and classification in sensor networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 691–706, 2011.

[21] C. Czado, E. C. Brechmann, and L. Gruber, "Selection of vine copulas," in *Copulae in Mathematical and Quantitative Finance*. Springer, 2013, pp. 17–37.

[22] C. Czado, S. Jeske, and M. Hofmann, "Selection strategies for regular vine copulae," *Journal de la Société Française de Statistique*, vol. 154, no. 1, pp. 174–191, 2013.

[23] T. Damarla, "Hidden markov model as a framework for situational awareness," in *Proc. 11th International Conference on Information Fusion*, 2008, pp. 1–7.

[24] M. Davy and A. Doucet, "Copulas: a new insight into positive time-frequency distributions," *IEEE Signal Processing Letters*, vol. 10, no. 7, pp. 215–218, Jul. 2003.

[25] J. Deng, Y. S. Han, W. B. Heinzelman, and P. K. Varshney, "Balanced-energy sleep scheduling scheme for high-density cluster-based sensor networks," *Computer communications*, vol. 28, no. 14, pp. 1631–1642, 2005.

[26] ——, "Scheduling sleeping nodes in high density cluster-based sensor networks," *Mobile Networks and Applications*, vol. 10, no. 6, pp. 825–835, 2005.

[27] J. Dissmann, E. C. Brechmann, C. Czado, and D. Kurowicka, "Selecting and estimating regular vine copulae and application to financial returns," *Computational Statistics & Data Analysis*, vol. 59, pp. 52–69, 2013.

[28] E. Drakopoulos and C.-C. Lee, "Optimum multisensor fusion of correlated local decisions," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 27, no. 4, pp. 593–606, 1991.

[29] N. E. D. Elmadany, Y. He, and L. Guan, "Information fusion for human action recognition via biset/multiset globality locality preserving canonical correlation analysis," *IEEE Transactions on Image Processing*, vol. 27, no. 11, pp. 5275–5287, 2018.

[30] J. Fang and H. Li, "Distributed adaptive quantization for wireless sensor networks: From delta modulation to maximum likelihood," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5246–5257, 2008.

[31] ——, "Power constrained distributed estimation with cluster-based sensor collaboration," *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3822–3832, 2009.

[32] ——, "Power constrained distributed estimation with correlated sensor data," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3292–3297, 2009.

[33] J. Fang, H. Li, Z. Chen, and S. Li, "Optimal precoding design and power allocation for decentralized detection of deterministic signals," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 3149–3163, 2012.

[34] G. Fellouris and G. V. Moustakides, "Decentralized sequential hypothesis testing using asynchronous communication," *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 534–548, 2011.

[35] L. Gao, L. Qi, E. Chen, and L. Guan, "Discriminative multiple canonical correlation analysis for information fusion," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1951–1965, 2017.

[36] C. Genest, B. Rémillard, and D. Beaudoin, "Goodness-of-fit tests for copulas: A review and a power study," *Insurance: Mathematics and economics*, vol. 44, no. 2, pp. 199–213, 2009.

[37] H. Guo, L. Chen, L. Peng, and G. Chen, "Wearable sensor based multimodal human activity recognition exploiting the diversity of classifier ensemble," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 1112–1123.

[38] W. Guo, J. Wang, and S. Wang, "Deep multimodal representation learning: A survey," *IEEE Access*, vol. 7, pp. 63 373–63 394, 2019.

[39] M. Haghighat, M. Abdel-Mottaleb, and W. Alhalabi, "Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 1984–1996, 2016.

[40] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "A non-reference image fusion metric based on mutual information of image features," *Computers & Electrical Engineering*, vol. 37, no. 5, pp. 744–756, 2011.

[41] H. He, "Heterogeneous sensor signal processing for inference with nonlinear dependence," Ph.D. dissertation, Syracuse University, 2015.

[42] ——, "Heterogeneous sensor signal processing for inference with nonlinear dependence," Ph.D. dissertation, Syracuse University, Syracuse, NY, December 2015.

[43] H. He, A. Subramanian, S. Choi, P. K. Varshney, and T. Damarla, "Social media data assisted inference with application to stock prediction," in *2015 49th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2015, pp. 1801–1805.

[44] H. He, A. Subramanian, X. Shen, and P. K. Varshney, "A coalitional game for distributed estimation in wireless sensor networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 4574–4578.

[45] H. He, A. Subramanian, P. K. Varshney, and T. Damarla, "Fusing heterogeneous data for detection under non-stationary dependence," in *2012 15th International Conference on Information Fusion (FUSION)*. IEEE, 2012, pp. 1792–1799.

[46] H. He and P. K. Varshney, "A coalitional game for distributed inference in sensor networks with dependent observations," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1854–1866, 2016.

[47] A. M. Hussain, "Multisensor distributed sequential detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 3, pp. 698–708, 1994.

[48] S. G. Iyengar, "Decision making with heterogeneous sensors - a copula based approach," Ph.D. dissertation, Syracuse University, Syracuse, NY, August 2011.

[49] S. G. Iyengar, P. K. Varshney, and T. Damarla, "On the detection of footsteps based on acoustic and seismic sensing," in *2007 41st Asilomar Conference on Signals, Systems and Computers*, 2007.

[50] ——, "A parametric copula-based framework for hypothesis testing using heterogeneous data," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2308–2319, 2011.

[51] H. Joe, "Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters," *Lecture Notes-Monograph Series*, pp. 120–141, 1996.

[52] J. Joung, E. Kurniawan, and S. Sun, "Channel correlation modeling and its application to massive mimo channel feedback reduction," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 3787–3797, 2016.

[53] M. Kafai and B. Bhanu, "Dynamic Bayesian Networks for Vehicle Classification in Video," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 100–109, 2012.

[54] M. Kam, Q. Zhu, and W. S. Gray, "Optimal data fusion of correlated local decisions in multiple sensor detection systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 3, pp. 916–920, 1992.

[55] S. Kar and J. M. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1383–1400, 2010.

[56] S. Kar and P. K. Varshney, "Linear coherent estimation with spatial collaboration," *IEEE Transactions on Information Theory*, vol. 59, no. 6, pp. 3532–3553, 2013.

[57] H. Kasasbeh, L. Cao, and R. Viswanathan, "Hard decision based distributed detection in multi-sensor system over noise correlated sensing channels," in *Information Science and Systems (CISS), 2016 Annual Conference on*. IEEE, 2016, pp. 280–285.

[58] ——, "Soft-decision based distributed detection over correlated sensing channels," in *Information Sciences and Systems (CISS), 2017 51st Annual Conference on*. IEEE, 2017, pp. 1–6.

[59] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, New Jersey: Prentice-Hall Inc., 1998.

[60] P. Khanduri, D. Pastor, V. K. Sharma, and P. K. Varshney, "on random distortion testing based sequential non parametric hypothesis testing," in *56th annual Allerton Conference on communication, control and computing*, 2018.

[61] A. Krasnopeev, J.-J. Xiao, and Z.-Q. Luo, "Minimum energy decentralized estimation in a wireless sensor network with correlated sensor noises," *EURASIP Journal on Wireless Communications and Networking*, vol. 2005, no. 4, pp. 473–482, 2005.

[62] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.

[63] J. Li and P. Stoica, "Mimo radar with colocated antennas," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 106–114, 2007.

[64] M. Li, V. Rozgić, G. Thatte, S. Lee, A. Emken, M. Annavaram, U. Mitra, D. Spruijt-Metz, and S. Narayanan, "Multimodal physical activity recognition by fusing temporal and cepstral information," *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 18, no. 4, p. 369, 2010.

[65] S. Li, X. Li, X. Wang, and J. Liu, "Decentralized sequential composite hypothesis test based on one-bit communication," *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 3405–3424, 2017.

[66] S. Liu, S. P. Chepuri, M. Fardad, E. Maşazade, G. Leus, and P. K. Varshney, "Sensor selection for estimation with correlated measurement noise," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3509–3522, 2016.

[67] S. Liu, S. Kar, M. Fardad, and P. K. Varshney, "Sparsity-aware sensor collaboration for linear coherent estimation," *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2582–2596, 2014.

[68] Z.-Q. Luo *et al.*, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE Journal on selected areas in communications*, vol. 23, no. 4, pp. 735–744, 2005.

[69] Y. Mei, "Asymptotic optimality theory for decentralized sequential hypothesis testing in sensor networks," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2072–2089, 2008.

[70] G. Mercier, G. Moser, and S. Serpico, "Conditional copula for change detection on heterogeneous sar data," in *Proc. IEEE International Geoscience and Remote Sensing Symposium IGARSS 2007*, Jul. 23–28, 2007, pp. 2394–2397.

[71] A. Mokh, M. Crussière, and M. Hélard, "Performance analysis of extended rask under imperfect channel estimation and antenna correlation," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2018, pp. 1–6.

[72] O. Morales Nápoles, *Bayesian belief nets and vines in aviation safety and other applications*. TU Delft, Delft University of Technology, 2010.

[73] O. Morales Napoles, R. M. Cooke, and D. Kurowicka, "About the number of vines and regular vines on n nodes," 2010.

[74] F. Moya Rueda, R. Grzeszick, G. A. Fink, S. Feldhorst, and M. ten Hompel, "Convolutional neural networks for human activity recognition using body-worn sensors," in *Informatics*, vol. 5, no. 2. Multidisciplinary Digital Publishing Institute, 2018, p. 26.

[75] R. B. Nelsen, *An introduction to copulas*. Springer Science & Business Media, 2013, vol. 139.

[76] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 689–696.

[77] X. Nguyen, M. J. Wainwright, and M. I. Jordan, "On optimal quantization rules for sequential decision problems," in *Proc. IEEE Int. Symp. Inf. Theory*, 2006, pp. 2652–2656.

[78] R. Niu and P. K. Varshney, "Target Location Estimation in Sensor Networks With Quantized Data," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4519–4528, 2006.

[79] H. F. Nweke, Y. W. Teh, G. Mujtaba, and M. A. Al-garadi, "Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions," *Information Fusion*, vol. 46, pp. 147–170, 2019.

[80] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[81] C. Papatsimpa and J.-P. M. Linnartz, "Energy efficient communication in smart building wsn running distributed hidden markov chain presence detection algorithm," in *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*. IEEE, 2018, pp. 112–117.

[82] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002.

[83] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 8, pp. 1226–1238, 2005.

[84] V. Radu, C. Tong, S. Bhattacharya, N. D. Lane, C. Mascolo, M. K. Marina, and F. Kawsar, "Multimodal deep learning for activity and context recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, p. 157, 2018.

[85] R. Raghavendra, S. Venkatesh, K. B. Raja, F. A. Cheikh, and C. Busch, "Mutual information based multispectral image fusion for improved face recognition," in *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. IEEE, 2016, pp. 62–68.

[86] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 96–108, 2017.

[87] I. D. Schizas, G. Mateos, and G. B. Giannakis, "Distributed lms for consensus-based in-network adaptive processing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2365–2382, 2009.

[88] G. Schwarz *et al.*, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[89] M. Stikic, T. Huynh, K. Van Laerhoven, and B. Schiele, "Adl recognition based on the combination of rfid and accelerometer sensing," in *Pervasive Computing Technologies for Healthcare, 2008. PervasiveHealth 2008. Second International Conference on*. IEEE, 2008, pp. 258–263.

[90] A. Stisen, H. Blunck, S. Bhattacharya, T. S. Prentow, M. B. Kjærgaard, A. Dey, T. Sonne, and M. M. Jensen, "Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition," in *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2015, pp. 127–140.

[91] Storm, http://storm-project.net/, 2016.

[92] A. Subramanian, A. Sundaresan, and P. K. Varshney, "Fusion for the detection of dependent signals using multivariate copulas," in *2011 Proceedings of the 14th International Conference on Information Fusion (FUSION)*. IEEE, 2011, pp. 1–8.

[93] A. Sundaresan, "Detection and location estimation of a random signal source using sensor networks," Ph.D. dissertation, Syracuse University, Syracuse, NY, December 2010.

[94] A. Sundaresan and P. K. Varshney, "Location estimation of a random signal source based on correlated sensor observations," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 787–799, 2011.

[95] A. Sundaresan, P. K. Varshney, and N. S. Rao, "Copula-based fusion of correlated decisions," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 1, pp. 454–471, 2011.

[96] A. Tamjidi, R. Oftadeh, S. Chakravorty, and D. Shell, "Efficient distributed state estimation of hidden markov models over unreliable networks," in *2017 International Symposium on Multi-Robot and Multi-Agent Systems (MRS)*. IEEE, 2017, pp. 112–119.

[97] S. Tantaratana and J. Thomas, "Relative efficiency of the sequential probability ratio test in signal detection," *IEEE Transactions on Information Theory*, vol. 24, no. 1, pp. 22–31, 1978.

[98] S. Tantaratana and J. B. Thomas, "Truncated sequential probability ratio test," *Information Sciences*, vol. 13, no. 3, pp. 283–300, 1977.

[99] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential analysis: Hypothesis testing and changepoint detection*. Chapman and Hall/CRC, 2014.

[100] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Transactions on Aerospace and Electronic systems*, no. 4, pp. 501–510, 1981.

[101] H. L. V. Trees, *Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV)*. Wiley-Interscience, 2002.

[102] J. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 440–446, 1985.

[103] J. N. Tsitsiklis, "Extremal properties of likelihood-ratio quantizers," *IEEE Transactions on Communications*, vol. 41, no. 4, pp. 550–558, 1993.

[104] P. K. Varshney, "Distributed bayesian detection: Parallel fusion network," in *Distributed Detection and Data Fusion*. Springer, 1997, pp. 36–118.

[105] ——, *Distributed detection and data fusion*. Springer Science & Business Media, 2012.

[106] V. V. Veeravalli, T. Basar, and H. V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 433–442, 1993.

[107] A. Wald, "Sequential tests of statistical hypotheses," *The annals of mathematical statistics*, vol. 16, no. 2, pp. 117–186, 1945.

[108] L. Wassermann, "All of nonparametric statistics," 2006.

[109] P. Willett, P. F. Swaszek, and R. S. Blum, "The good, bad and ugly: distributed detection of a known signal in dependent gaussian noise," *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3266–3279, 2000.

[110] P. Wu, H.-K. Peng, J. Zhu, and Y. Zhang, "Senscare: Semi-automatic activity summarization system for elderly care," in *International Conference on Mobile Computing, Applications, and Services*. Springer, 2011, pp. 1–19.

[111] P. Wu, J. Zhu, and J. Y. Zhang, "Mobisens: A versatile mobile sensing platform for real-world applications," *Mobile Networks and Applications*, vol. 18, no. 1, pp. 60–80, 2013.

[112] Y. Xin, H. Zhang, and L. Lai, "A low-complexity sequential spectrum sensing algorithm for cognitive radio," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 3, pp. 387–399, 2014.

[113] X. Xiong, K. L. Chan, and K. L. Tan, "Similarity-driven cluster merging method for unsupervised fuzzy clustering," in *Proceedings of the 20th conference on Uncertainty in artificial intelligence*. AUAI Press, 2004, pp. 611–618.

[114] J. Yang, M. N. Nguyen, P. P. San, X. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition." in *Ijcai*, vol. 15, 2015, pp. 3995–4001.

[115] H. Yoon and H. Sompolinsky, "The effect of correlations on the fisher information of population codes," in *Advances in neural information processing systems*, 1999, pp. 167–173.

[116] P. Zappi, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Troster, "Activity recognition from on-body sensors by classifier fusion: sensor scalability and robustness," in *Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on*. IEEE, 2007, pp. 281–286.

[117] S. Zhang, S. Liu, V. Sharma, and P. K. Varshney, "Optimal sensor collaboration for parameter tracking using energy harvesting sensors," *IEEE Transactions on Signal Processing*, vol. 66, no. 12, pp. 3339–3353, 2018.

VITA

NAME OF AUTHOR: Shan Zhang

PLACE OF BIRTH: Yicheng, Shanxi, China

DATE OF BIRTH: September 22, 1990

EDUCATION: B.S. June 2014 University of Science & Technology Beijing, China

PROFESSIONAL EXPERIENCE:

- Graduate Teaching Assistant, Department of Electrical Engineering and Computer Science, Syracuse University, August 2014 - May 2015
- Graduate Research Assistant, Sensor Fusion Laboratory, Department of Electrical Engineering and Computer Science, Syracuse University, May 2015 - December 2019

PUBLICATIONS:

1. Shan Zhang, Jielong Xu, Sora Choi, Jian Tang, Pramod K. Varshney, and Zhenhua Chen, “A Parallel Platform for Fusion of Heterogeneous Stream Data," in *Proc. 21th International Conference on Information Fusion*, 2018.
2. Shan Zhang, Baocheng Geng, Pramod K. Varshney, and Muralidhar Rangaswamy, “Fusion of Deep Neural Networks for Activity Recognition: A Regular Vine Copula Based Approach," in *Proc. 22th International Conference on Information Fusion*, 2019.

3. Shan Zhang, Prashant Khanduri, and Pramod K. Varshney, “Distributed Sequential Hypothesis Testing with Dependent Sensor Observations," in *Proc. 53rd IEEE Asilomar Conference on Signals Systems and Computer*, 2019.

4. Shan Zhang, Lakshmi N. Theagarajan, Sora Choi, and Pramod K. Varshney, “Fusion of Correlated Decisions Using Regular Vine Copulas," *IEEE Transactions on Signal Processing*, vol 67, no. 8, pp. 2066-2079, 2019.

5. Shan Zhang, Prashant Khanduri, and Pramod K. Varshney, “Distributed Sequential Detection: Dependent Observations and Imperfect Communication," *IEEE Transactions on Signal Processing*, accepted, 2019.

6. Shan Zhang, Pranay Sharma, and Pramod K. Varshney, “Distributed Estimation in Large Scale Wireless Sensor Networks via A Two-Step Cluster-based Approach," to be submitted, 2019.

7. Shan Zhang, Sijia Liu, Vinod Sharma, and Pramod K. Varshney, “Optimal Sensor Collaboration for Parameter Tracking Using Energy Harvesting Sensors," *IEEE Transactions on Signal Processing*, vol 66, no. 12, pp. 3339-3353, 2018.

8. Shan Zhang, Aditya Vempaty, Susan E. Parks, and Pramod K. Varshney, “On Classification of Environmental Acoustic Data Using Crowds," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2017.

9. Kush R. Varshney, Prashant Khanduri, Pranay Sharma, Shan Zhang, and Pramod K. Varshney, “Why Interpretability in Machine Learning? An Answer Using Distributed Detection and Data Fusion Theory," in *ICML Workshop on Human Interpretability in Machine Learning (WHI)*, 2018.